Tell Me Something I Don't Already Know: Informedness and the Impact of Information Programs

David P. Byrne* Andrea La Nauze† Leslie A. Martin*

June 22, 2017

Abstract

We document how imperfect information generates heterogeneous effects in information treatments with personalized high-frequency feedback and peer comparisons. In our field experiment in retail electricity, we find that high and low energy users symmetrically underestimate and overestimate their relative energy use pre-treatment. Responses to personalized feedback, however, are asymmetric. Households that overestimate their relative use and low users both respond by consuming more. These boomerang effects provide evidence that peer-comparison information programs, even those coupled with normative comparisons, are not guaranteed to lead to increases in prosocial behavior.

Keywords: Imperfect information, high-frequency feedback, peer comparison, learning, field experiment, retail electricity

JEL Codes: C93, D12, D83, L94, Q41

---

*Department of Economics, The University of Melbourne, 111 Barry Street, Parkville, VIC 3010, Australia. †Department of Economics, The University of Pittsburgh, Wesley Posvar Hall, Pittsburgh, PA 15260. Email addresses: `byrned@unimelb.edu.au`, `lanauze@pitt.edu`, and `leslie.martin@unimelb.edu.au`. Three anonymous referees have provided comments that substantially improved the paper. We also thank Victoria Baranov, Yann Burden, Koichiro Ito, Lorenz Goette, Jia Sheen Nah, David Rapson, Tom Wilkening, and many workshop participants. Funding from the Australian Research Council, University of Melbourne, and Institute for a Broadband Enabled Society is appreciated. All errors, omissions, and views are our own.

1 Introduction

Information programs are increasingly being used to encourage social and behavioral change. For instance, the formation of the Social and Behavioral Sciences Team in 2014 by the Obama Administration put non-pecuniary, nudge-based information treatments at the forefront of U.S. public policy. Interventions that provide individuals with peer comparisons have enhanced voter turnout (Gerber and Rogers 2009), rates of charitable giving (Frey and Meier 2004), retirement savings (Beshears et al. 2015), water conservation (Ferraro and Price 2013, Ferraro and Miranda 2013, Bernedo, Ferraro and Price 2014, and Brent et al. 2015) and energy conservation (Allcott 2011 and Costa and Kahn 2013). These and other information treatments are popular because they are low-cost alternatives to politically-sensitive regulation, and because a growing body of research highlights their effectiveness in achieving policy goals.[1]

In this paper, we provide evidence that peer-comparison treatments should be treated with some caution. Peer comparisons are hypothesized to work by moving individuals towards peer-group averages. Schultz et al. 2007 warn that although some individuals will learn that they are failing to meet the norm, others will learn that they exceed it. Individuals that underestimate their relative levels of prosocial behavior will potentially *decrease* their engagement in such behavior if shown peer comparisons. Schultz et al. 2007 further note that systematic overconfidence about relative prosocial behavior reduces the overall impact of such potential *boomerang effects*.[2] Intuitively, peer comparisons have the greatest impact in generating prosocial behavior when those who exceed the norm already know that they do, while those who are failing to meet it do not.

[1]Interventions that solely provide information on the costs and benefits of individual behavior can reduce the prevalence of diarrhea (Kremer et al. 2011), increase hand-washing (Luby et al. 2005), reduce risky sexual behavior (Dupas 2011), encourage retirement savings plan participation (Duflo and Saez 2003), and reduce caloric intake (Bollinger et al. 2011).

[2]Boomerang effects have arisen from peer-comparison treatments in various domains: retirement savings enrollment (Beshears et al. 2015), tax compliance (Fellner et al. 2013), student academic performance (Carrell et al. 2013) and worker performance in training programs (Ashraf et al. 2014).

Although the literature on information treatments is extensive, few have documented the impact of pre-treatment informedness on treatment outcomes. This paper sheds light on this issue using a field experiment with a novel two-step research design. We first run a baseline survey that elicits individuals' beliefs about their relative levels of prosocial behavior. We then randomly inform individuals about how their behavior actually compares with that of their peers. Then we track their levels of prosocial behavior over time. To our knowledge, we are the first to document how imperfect information about peer behavior generates heterogeneous effects in information programs.

Our research context is retail electricity and the prosocial behavior of interest is energy conservation. Like previous interventions that we review below, our information treatment provides peer comparisons along with detailed information on energy use and energy-saving tips. Unlike previous studies, our treatment households receive this information through biweekly emails and a smart-meter web-portal with half-hourly updates on energy use. Previous interventions provide monthly or quarterly information through paper bills.

We focus on information programs in this setting for several reasons. First, behavioral interventions that inform households about energy conservation strategies and provide peer comparisons have been used extensively in electricity markets, yielding highly heterogeneous treatment effect estimates (Delmas et al. 2013). This provides *a priori* motivation for investigating the underlying channels of treatment effects. Moreover, boomerang effects are a first-order concern in electricity information programs intended to increase energy conservation.[3] Finally, high-frequency, high-quality data on household behavior are available to explore treatment effect heterogeneity and its determinants.

Our paper delivers three main results. First, we find no evidence that households systematically underestimate relative energy use. Low energy users are no more informed about their relative use than high users. In our sample of over 1000 households, just as many respondents overestimate their relative use as do underestimate. Most households think that they are "average" users, and high and low energy users symmetrically underestimate

[3]Organizations like OPower explicitly attempt to mitigate boomerang effects by simultaneously providing "injunctive norms," that is, reminders to all users that energy conservation is prosocial, and rewarding low users with smiley face emojis or other forms of praise.

and overestimate, respectively, their positions in the energy-use distribution. We check for overconfidence using a test derived from a Bayesian updating model from Burks et al. 2013 and find no evidence of overconfidence in our data.

The second set of results focuses on the extent to which pre-treatment informedness generates heterogeneous treatment effects from our information program. We allow treatment effects to simultaneously depend on pre-treatment energy use level, census block-level demographics, and environmentalism. We also account for other important factors such as potential differences in attrition between treatment and control households, mean reversion, and seasonality. We find a boomerang effect among households that underestimate their success at conserving energy. Specifically, households who overestimate their relative energy use subsequently increase their use by 6.3% in response to treatment relative to other households at similar levels of use. In contrast, and perhaps surprisingly, we find that those who underestimate their relative energy use do not exhibit a noticeable change in energy use in response to treatment. In sum, while we find symmetric errors-in-beliefs among similar numbers of households who overestimate and underestimate their relative energy use, we find asymmetric responses to treatment across these groups.

Going beyond our baseline estimates, we investigate how these heterogeneous treatment effects evolve over time. We find that the informational boomerang effect among overestimators dissipates after six months of treatment. This suggests a short-lived salience effect from becoming informed about peer behavior in our setting.

Our third set of results documents a second boomerang effect. Regardless of beliefs, households in the lowest quintile of baseline energy use increase their energy use by 11.7% in response to our information treatment, while households in the highest quintile decrease their use by 11.0%. In other words, after controlling for beliefs about relative use, low and high energy users both exhibit a reversion towards the mean following treatment above and beyond any regression towards the mean exhibited by control group peers with the same pre-treatment energy-use levels. Moreover, these energy use-based heterogeneous effects are sustained over our entire seven-month treatment period.

We offer a few explanations for this second boomerang effect. It could be due to household learning and reoptimization. For example, low-use households that learn through our information treatment that they previously overestimated the savings from certain behaviors (like switching off lights) may put less effort into regulating that behavior. Psychology may also play a role: Brehm's theory of psychological reactance predicts an opposition effect in response to information if households resent being told what to do.[4] Although we cannot rule out that an oppositional reaction drives the increase in post-treatment use among low users, the persistent nature of the effect suggests that households are reoptimizing. Making this distinction is important because reoptimizing suggests welfare improvements whereas oppositional reactions represent psychological costs. To the extent that boomerang effects come from social or individual learning, information treatments may be welfare-improving for households even if they increase overall electricity use.[5]

Related literature

Our main contribution is to the cited literature on information programs. Our study distinguishes itself by combining estimates of pre-treatment beliefs with an information treatment that aims to correct these beliefs. More specifically, we innovate on previous studies that either: (1) document biases in beliefs without attempting to directly change them[6] or (2) provide information without measuring baseline beliefs—or only measure beliefs with post-treatment surveys—and yet interpret treatment effects as arising from pre-treatment errors-in-beliefs.[7] Through our experimental design, we are able to directly study rational belief updating and show how pre-treatment errors-in-beliefs can lead to unintended consequences of information programs.

This paper also builds on prior studies of information interventions in electricity markets

[4]By emphasizing energy conservation, our information treatment could induce such a resentment-based increase in energy use among all households. Among high users, however, this effect may be dominated by a conservation effect from individuals learning about how to reduce electricity bills.

[5]Our analysis focuses on evaluating the effects of an information treatment on energy conservation; we do not undertake explicit welfare calculations. Although a common policy objective, energy conservation may not be welfare improving if, for example, prices are well-above long run marginal cost (Davis and Muehlegger 2010), or if the costs of conservation outweigh the social benefits of reduced externalities (Allcott and Kessler 2015).

[6]Chetty 2009, Jensen 2010, Bollinger et al. 2011, Kling et al. 2012, Burks et al. 2013

[7]For instance, Schultz et al. 2007, Jalan and Somanathan 2008, Dupas 2011, among many others.

in two ways. Our focus on the determinants of heterogeneous treatment effects most closely relate to Allcott 2011 and Costa and Kahn 2013. These authors provide evidence of how differences in energy use (e.g., high vs. low users) and political ideology (e.g., Liberals vs. Conservatives) predict heterogeneous effects from information programs. However neither are able to separate the role of imperfect information from other factors. Therefore, we contribute to the literature by providing a direct test of how imperfect information leads to heterogeneous treatment effects in electricity market information programs.

Moreover, we are the second paper to document a boomerang effect in electricity use, and the first to identify one while accounting for seasonality and mean reversion.[8] Why has a boomerang effect not been discovered previously? We offer two explanations. First, previous studies offered feedback at lower frequencies (e.g., monthly or quarterly) via paper bills. We provide information at a higher frequency, delivered via a technology-intensive medium: emails and web portals, usually accessed via smart phones. As a result our treatment provides greater opportunities to learn about the cost of personal energy use.[9]

Second, as Allcott 2015 documents, earlier electricity market information programs may have taken place in contexts particularly suited to generating conservation effects and not boomerang effects. Both of these explanations provide reason to believe that future treatments, which are likely to rely on high-frequency, smart-meter information and involve less specially-targeted populations, may also be more likely to exhibit boomerang effects.

---

[8]Schultz et al. 2007 document boomerang effects. Allcott 2011, Ayres et al. 2012, and Costa and Kahn 2013, who control for seasonality and mean reversion, as we do, do not find boomerang effects. Ferraro and Price 2013 and Ferraro and Miranda 2013 do not find a boomerang effect potentially because their experimental sample excludes the lowest quintile of users in the population.

[9]Alcott 2011 documents that monthly and bimonthly treatments generate 30% larger effects than quarterly treatments. Wichman 2017 and Goette et al. 2017 similarly find that providing higher-frequency feedback on personal use in residential water markets can generate large behavioral effects.

2 The Experiment

2.1 Context

The context for our study is the retail electricity market in the Australian state of Victoria. Approximately 70% of Victorians reside in Melbourne, a city with four million people. Although the climate is similar to that of San Francisco, Melbourne is known for having "four seasons in a day" with both warm and cold air masses from the Outback and Antarctic affecting local weather. The average retail residential electricity price in the state is 25 cents US per kWh. This price is high compared to the US (10 cents per kWh), but low compared to Europe (30 cents per kWh) (Mountain 2012).

The state has retail market competition. Households actively switch electricity retailers in pursuit of lower retail prices and better service, with one in five households changing their electricity retailer each year. Upstream electricity distributors, in contrast, are geographic monopolists who pass on regulated network charges to retailers, and ultimately households. There were five distributors and sixteen retailers in the market during our sample period.

Starting in June 2010, the market experienced major technological change with a government-mandated smart-electricity-meter rollout. Unlike traditional electricity metering which provides energy-use information at monthly or quarterly frequencies, these new meters collect high-frequency, half-hourly readings of household-level electricity use. The mandate required all upstream distributors to install smart meters for every residential household and small business in the state by December 2013. Retailers had no influence over where or when smart meters were provided to their households.[10]

Against this context, we partnered with an electricity retailer in 2012 to study the effects of providing personalized, high-frequency smart-meter information to households. The retailer was a small player in the market with a customer base of less than 30,000 households nationwide. To engage households with their smart-meter data, the retailer

[10]We study a market with a recent and all-encompassing smart-meter rollout. External validity concerns apply if considering existing or future markets with limited smart-meter penetration, or markets that have had smart meters for a long time.

employed a third-party web-portal developer called Billcap. Billcap provided households with interactive online home energy reports that provided their smart-meter data in a simple visual format, along with energy-saving tips, and peer comparisons.

Moreover, Billcap increased the frequency with which households engaged with information on electricity use and costs. Specifically, their platform sent biweekly emails that summarized households' electricity use and costs over the previous two weeks. The emails also prompted households to engage with their online reports. Figure 1 presents a typical home energy report; Billcap's emails presented similar information to that in panel (b). Prior to our experiment, households' main source of information on electricity use was their quarterly bills. The Billcap information treatment was thus expected to create large changes in households' information sets regarding the relationship between daily energy use and costs, and peer behavior.[11]

2.2 Research design

Baseline survey

Prior to treatment, we ran a baseline survey via email and a linked website to collect individual-level data on home characteristics such as the number of rooms and residents, gas appliances, and air conditioning. A key innovation of our experimental design is that we also elicited households' beliefs about their relative energy use with the following question:

> *Compared to energy use in Melbourne homes as large as yours, what statement best describes your household's monthly energy use?*
> a. *High (top 20%)*
> b. *Above average (top 40%)*
> c. *Average*
> d. *Below average (bottom 40%)*
> e. *Low (bottom 20%)*

Using the beliefs data, we compared the quintile of the use distribution that a household believed it was in to the quintile it was actually in, conditional on the number of rooms

[11]The retailer was the first in the country to use Billcap's platform. Households were unaware of Billcap and its services prior to October 2012. At the time of Billcap's introduction only one other electricity retailer in the market (AGL) offered a smart-meter web portal.

in their home.[12] With this comparison we identify whether a household overestimated or underestimated its energy use relative to its peers prior to treatment.

There are two notable points with the beliefs question. We intentionally included a middle, neutral response so we could distinguish households who commit to identifying themselves as being above or below average use. We also left the size comparisons vague (e.g., "as large as") to simplify the question and mitigate a perceived risk of household frustration by our partner retailer.[13]

We invited survey responses from 8564 residential households without solar panels whose energy use did not exceed 50 kWh/day. In total, 1188 households responded to the survey (14% response rate).

Information treatment

We used a phase-in design to evaluate the effect of our information treatment on residential energy use. We restricted our sample to the 2423 households with smart-meter readings dating back to at least October 1, 2012. We then provided a randomly chosen subset of households with access to Billcap's biweekly emails and home energy reports in two treatment waves: October 2012 (640 households) and March 2013 (853 households). For technical reasons relating to the Billcap platform, we were not able to randomize based on survey response. In this restricted sample, 311 answered the survey (13% response rate). The 930 control group households were not eligible for treatment until June 2013. At this time, the retailer rolled out the Billcap service company-wide, which put an end to our experiment. Importantly, only treatment households received emails and could access the personalized smart-meter web portal. Control households were not sent emails and were not able to access the web portal.

[12] Our results are unchanged if we instead condition on the number of residents in the home.

[13] We emailed the survey to households at most twice, and included a range of questions on household-level characteristics for our empirical analysis. This helped mitigate the risk that the act of responding to baseline surveys could lead respondents with different pre-treatment beliefs to change their consumption in response to treatment in systematically-biased ways. See Zwane et al. 2011 for evidence on how baseline surveys can, in some cases, modify post-treatment behavior in randomized control trials.

3 Data

In this section we discuss our data sources and present some summary statistics. We also investigate two potential sources of sample selection bias in evaluating the effect from our information treatment: survey response bias and attrition.

3.1 Data sources and summary statistics

From our partner electricity retailer we obtain half-hourly smart-meter data on energy use, as well as the exact dates each household starts and ends their electricity contract with the retailer. We use these data to track households' energy use and retailer switching decisions from July 1, 2012 to May 30, 2013.[14]

We combine the electricity-meter data with data on household characteristics. Our baseline survey provides us with household-level data on number of bedrooms, occupancy, air-conditioning, and beliefs about relative energy use. In addition, we collect census block-level demographic information from the Australia Bureau of Statistics and match it to households in our sample. The lowest level of aggregation available is at the "Statistical Area Level" (SA1) census tract, which typically contains 250 homes. Motivated by Costa and Kahn 2013, we also collect elections data on the local shares of votes for the federal Green Party. These data potentially capture each electoral district's revealed preferences for pro-environmental information programs such as ours.

Tables 1 and 2 present summary statistics for pre-treatment electricity use and all survey and demographic data. Table 1 shows that mean daily pre-treatment energy use in the October 2012, March 2013, and June 2013 treatment waves share common trends and do not exhibit statistically significant differences.[15] Table 2 shows some significant differences

[14] Billcap collects web-use data that allows us to see whether a household opens an email. 47% of all treatment emails sent are viewed by households. More than 80% households viewed emails using smart phones. These data allow us to estimate a Local Average Treatment Effect (LATE) of households engaging with the information in their emails and home energy reports.

[15] The first rows combine the data for the March 2013 and June 2013 waves because prior to March 2013 they act as a single control group for the first wave of treatment. The June 2013 treatment wave

in household characteristics and census block-level demographics across waves. Households in the October 2012 treatment wave are more likely to live in freestanding houses with more rooms than households in the March and June 2013 waves. In addition, households in the June 2013 treatment wave have lower incomes, are younger, and are less likely to be employed full-time than households in the other waves. These differences motivate our use of household fixed effects and post-treatment trends interacted with household characteristics and demographics in our regressions below.

3.2 Sample selectivity

Survey response bias

Survey sample selection bias may be an important confound in identifying how treatment effects differ depending on survey-based prior beliefs on relative energy use. However, we find no evidence of systematic survey response bias. Table 3 presents summary statistics for the subset of households with smart-meter readings dating back to at least October 1, 2012 by survey response status. Although survey respondents have slightly higher incomes and are older, respondents and non-respondents have similar pre-treatment energy use.

The Appendix provides further evidence of the lack of systematic differences between survey respondents and non-respondents. Figures A.1 and A.2 show that there are no statistically significant differences in the distributions of monthly and daily pre-treatment electricity use for the two groups. Table A.2 shows that there are no significant differences in observables among survey respondents and non-respondents in the full sample of households who were emailed the baseline survey.

Attrition

The high degree of retailer switching in the market raises attrition as another potential source of sample selection bias, as we are unable to observe households' electricity use once they switch retailers. Indeed, Table 3 reveals that households who switch away from our partner retailer at some point between October 2012 and June 2013 are different from those that remain. Switchers consume more electricity, have higher incomes, are more likely to be

corresponds to the end of our treatment due to the company-wide rollout of Billcap.

employed full-time, and live in areas where the federal Green Party gets more votes. Figures A.3 and A.4 in the Appendix show that the distributions of pre-treatment energy use among switchers and non-switchers further exhibit some statistically-significant differences.

More importantly, attrition rates differ among treatment and control households. This difference can be seen by comparing the five-month attrition rate during the October 2012 to February 2013 period for the October treatment wave to the attrition rate for the control group during this period. Of the 1712 households in the control group in this period, 539 (31.5%) switch away from the retailer, whereas only 120 (17.3%) of the 575 treatment households switch.

To formally test for differential attrition between treatment and control households, we estimate a logit model for a dummy variable that equals one if a household switches between October 2012 and February 2013, and zero otherwise. Our explanatory variables include an October 2012 treatment dummy, a baseline survey response dummy, average daily pre-treatment energy use and its square, and all the census demographic variables in Table 2. Our model also includes a dummy that equals one if a household has been with the retailer n months as of October 1, 2012, and zero otherwise. In total, we include 18 such dummies for $n = 1, \ldots, 18$. The estimation results imply treatment reduces the five-month attrition rate by 15.2 percentage points (s.e.=2.22, clustered at the household level), which is 50% of the 31.5 percentage point attrition rate for control households during this period.

We further find that although treatment reduces attrition, it does so in a way that is uncorrelated with pre-treatment energy use levels, beliefs about relative energy use, or demographics. The main predictors of attrition, other than treatment, are the month of the year and the number of months that a customer has been with the company.

Correcting for attrition bias

Although we find that attrition does not vary systematically among households with differing pre-treatment observable characteristics, our preferred specification corrects for differences between our treatment and control groups caused by differential attrition using inverse probability weights (IPWs). Specifically, we first estimate the propensity score $\hat{p}_{it}$ that household i in month t is in the treatment group. For month t, we estimate a logit model for a binary outcome variable that equals one if household i is in the treatment group in that month. The explanatory variables in the model are exactly those from the attrition logit model just discussed. We estimate separate logit models for each month of our experimental sample and use these models to predict $\hat{p}_{it}$. With these predictions, we reweigh treatment observations in month t by $1/\hat{p}_{it}$ and control observations by $1/(1-\hat{p}_{it})$.

Throughout, we report pairs cluster bootstrap standard errors and confidence intervals, where we cluster at the household level (e.g., the level of randomization of our treatment). In total we have 2423 clusters. In this way, we account for potential household-level shocks and estimation error in implementing the IPWs.[16]

Although our preferred specification involves IPWs, we note that our IPW-based results are very similar to results that apply no attrition correction (shown in Table A.5 in the Appendix and discussed in the results section below), and results that correct for attrition using Heckman's 1979 control function approach (available upon request).

4 Do households systematically underestimate their relative energy use?

In this section we present results from our pre-treatment survey. We look for evidence that some individuals learn they are failing to adhere to a norm through our information treatment, while others learn they exceed it.

There are several reasons to suspect that households would systematically underestimate their energy use relative to their peers. Prior research finds that individuals either underestimate their own participation in undesirable behaviors, or overestimate the extent of those behaviors among their peers (Kruger 1999 and Schultz et al. 2007). More broadly, there is a wide range of evidence from economics and psychology suggesting that individuals tend

[16]The standard errors we report are the same as those from Bitler et al. 2006. Like us, these authors use IPWs to correct for attrition bias in their estimates of welfare reform on labor market earnings and income arising from differential attrition between treatment and control households in their labor market experiment. We have checked and verified that our results are robust to clustering at the SA1 census-block level, or using the standard Moulton correction that does not account for IPWs estimation error.

to exhibit overconfidence, or "illusory superiority."[17] With respect to energy use, overconfidence may translate to a tendency to underestimate one's own energy use and overestimate that of their peers. Attari et al. 2010 document that individuals systematically underestimate the electricity used by most of their household appliances, especially larger ones. In a more recent survey, Opower 2014, documents that most households believe they make more of an effort and are more successful at conserving energy than others.

4.1 Descriptive statistics

Table 4 shows that, surprisingly, households in our sample do not systematically underestimate their energy use relative to their peers. The table presents the joint distribution of households' beliefs about their pre-treatment energy-use quintile and their actual quintile, conditional on the number of rooms in the home. Panel (a) contains the full sample of survey responses and panel (b) presents the sub-sample of households with smart-meter readings prior to October 1, 2012 (our experimental sample for the information treatment). If all households held correct beliefs, each cell along the diagonal would be 20% and all other cells would be empty. Panel (a) shows that only 25% of survey respondents are correct. More than half (58.06%) believe their energy use lies in the 40–60% quintile. Simply put, most people believe they are "average".

Among the households with incorrect beliefs in panel (a) there is an equal split between those who overestimate and underestimate their relative energy use. Figure 2 depicts this symmetry in prediction errors among high and low energy users. Panels (a) and (b) illustrate the prediction errors among all households, and for a sub-sample of households who believe they are above or below the middle quintile. Panel (c) presents the conditional distribution of prediction errors among households in the 20–40% and 60–80% actual use quintiles. In both groups, more than half of the households believe their energy use is in the 40–60% quintile. A similar pattern emerges in panel (d) where the prediction errors among households in the 1–20% and 80–100% quintiles are further accentuated.

[17]See Svenson 1981 (driving ability), Gilovich 2008 (leadership ability), Williams and Gilovich 2008 (intelligence, creativity, maturity, and positivity), DellaVigna and Malmendier 2006 (ability to commit to exercise) and Dunning et al. 2004 for an overview of the extensive literature on overconfidence.

When viewed from a slightly different perspective, the data reveal that relatively high-use and low-use households, on average, do better than a coin-flip in predicting whether they are relatively high or low use. And they do so symmetrically. Continuing to focus on the larger sample in panel (a) of Table 4, we see 59% of households at or above the 60^{th} percentile believe they are at or above the 60^{th} percentile. Similarly, 57% of households at or below the 40^{th} percentile believe they are at or below the 40^{th} percentile.

4.2 Testing for Bayesian updating

Going beyond simple descriptives, we can formally test whether households' beliefs and their actual relative electricity-use are consistent with overconfidence. Specifically, we implement a test proposed by Burks et al. 2013. This test looks at a joint distribution of beliefs and actual behavior at a given point in time and asks whether that joint distribution is consistent with an unbiased reaction to a history of noisy signals that contain true information about relative performance.

More specifically, the test derives from a model of belief formation where agents (households) are imperfectly informed about their relative performance (energy use compared to peers' use) and engage in Bayesian updating from a common prior as new information arrives from an arbitrarily-defined signal structure (e.g., historical bills, discussions with neighbors about electricity bills). If households truthfully report their posterior beliefs in our baseline survey, Burks et al. 2013 show that for any signal structure this Bayesian model has a testable implication for the joint distribution of beliefs and actual behavior: of all the households who believe they are in quantile k of the energy-use distribution, the largest (modal) share of them must actually be from quantile k. This implication of the model can be directly tested using our data on households' beliefs and their actual behavior.

The test is based on an empirical *allocation function*, $q_k(s_j)$, that returns the share of households in electricity-use quintile k who believe their quintile is j, where $\sum_j q_k(s_j) = 1$. Panel (a) of Table 5 produces the empirical allocation function across all quintiles for the full sample of 1188 survey respondents. The testable implication of the Bayesian model we explore is $q_k(s_k) = \max_\ell q_\ell(s_k)$ for all k; Burks et al. 2013 call this the *diagonal condition* since it implies the largest values for a given belief quintile should be entries along the diagonal in Table 5. Panel (a) of Table 5 shows the diagonal condition is in fact satisfied for three of the five belief quintiles.

The test is implemented in three steps. First, we find the vector $q = (q_k(s_j))_{k,j}$ that solves the following constrained maximum likelihood (ML) problem:

$$\max_q \sum_{j,k} n_{k,j} \log(q_k(s_j))$$
$$\text{subject to } 0 \leq q_k(s_j) \leq 1; \quad \sum_j q_k(s_j) = 1; \quad q_k(s_k) = \max_\ell q_\ell(s_k) \;\forall\; k \tag{1}$$

That is, we find the elements of the allocation function that simultaneously best fit the data while also satisfying the diagonal condition implied by the Bayesian model. Intuitively, if a given set of beliefs and actual behavior heavily violate the diagonal condition, the ML estimate of the allocation function and its empirical analogue will diverge, making it less likely the data are generated by the Bayesian model.

In the second step, we use the ML estimate of the allocation function to compute the following test statistic which quantifies the difference between the empirical allocation function and the theoretically-consistent ML estimate:

$$\hat{d} = \frac{1}{25}\sqrt{\sum_{j,k}(q_k(s_j) - q_k^{ML}(s_j))^2}$$

Finally, we construct a bootstrap distribution for $\hat{d}$ by simulating samples with our q^{ML} parameter estimates as per Burks et al. 2013. With this bootstrap distribution we can test the null that the empirical allocation function is generated by the Bayesian model.

Panel (b) of Table 5 produces the constrained ML estimate of the allocation function. The estimate is quite similar to its empirical analogue, which is perhaps not surprising given we find the diagonal condition is nearly satisfied empirically. The corresponding test statistic of $\hat{d} = 0.86$ has a bootstrap p-value of $p = 0.97$, implying that we cannot reject the null that households' beliefs and actual relative energy use is generated by the Bayesian model.[18] We find no evidence of systematic overconfidence in pre-treatment beliefs that

[18]Notice that this test is general in the sense that it applies for an arbitrary signal structure that households form their beliefs upon. The systematic individual-level prediction errors suggest this signal structure is quite noisy in our setting. However, given this signal structure, we find households' joint

would otherwise cause us to reject the null.

5 Informedness and the information program's effect

We now investigate how information affects energy use by interacting our information treatment with individual-level pre-treatment errors in beliefs about relative energy use. We hypothesize that households who overestimate (underestimate) their relative energy use will increase (decrease) their use once they become informed about their peers' behavior. In other words, we expect that correcting errors-in-beliefs will cause households to move towards becoming the users that they previously thought they were. Two prominent mechanisms emphasized in related literature on informational nudges[19] yield these predictions. The Becker 1965 household production model posits that households update their beliefs regarding the household utility-maximizing level of electricity use upon learning of their peers' behavior. The Levitt and List 2007 moral cost model focuses on changes in the perceived moral cost associated with the societal impact of their energy use on the environment. Both of these mechanisms rely on households reacting to new information.

Our analysis is developed over four parts, starting with suggestive graphical evidence that both pre-treatment beliefs and energy-use levels are important determinants of treatment effects. We then present nonparametric tests for the presence of treatment effect heterogeneity. Motivated by the graphs and these tests, we use regressions to investigate the sources of treatment effect heterogeneity, and the extent to which treatment effects persist over time.

5.1 Graphical evidence

Panels (a)–(f) of Figure 3 plot average daily electricity use for treatment and control households between July 2012 and May 2013 before and after treatment in event time. To create these figures, we remove household and month of year fixed effects, and plot the corresponding difference in electricity use between treatment and control households in event

distribution of beliefs and actual behavior is consistent with Bayesian updating.

[19]Allcott 2011, Costa and Kahn 2013, Ferraro and Price 2013

time for months $\tau = -2, -1, 0, 1, \ldots, 7$ before and after treatment.[20] Importantly, each panel reveals that there are no statistically or economically-significant pre-treatment differences in energy use between treatment and control households. The panels also highlight common pre-treatment trends in energy use between treatment and control within each subgroup.[21]

Panels (b)–(f) show that panel (a) masks heterogeneous treatment responses as a function of baseline energy use. Following treatment, treatment households in the lowest quintile of the pre-treatment use distribution (panel (b)) increase their energy use relative to other low-use control households who would otherwise have similar monthly trends in energy use. This graphical evidence is a first indicator that the information treatment causes an unintended boomerang effect. In contrast, treatment households in the highest quintile in panel (f) decrease their energy use relative to other high-use control households after treatment. The information treatment appears to have its intended conservation effect among this subgroup.

As per our discussion above, we ask: can this heterogeneity in program effects be explained by the fact that low users tend to overestimate their relative energy use, while high users tend to underestimate it? Figure 4 sheds some initial light on this question. Panel (a) plots daily average energy use, in event time removing household and month of year

[20]In estimating pre-treatment differences in energy use between treatment and control groups for months $\tau = -2, -1, 0$ we use monthly energy use for July, August and September 2012 for all households in the sample (e.g., the three months before the first treatment wave in October 2012), as well as household monthly energy use for December 2012, January and February 2013 (e.g., the three months before the second treatment wave in March 2013) for households who were not treated in the first treatment wave *and* who did not attrit between December 2012 and March 2013. For consistency, and to avoid introducing attrition bias into our estimates, we similarly focus on non-attritors between December 2012 and February 2013 in estimating pre-treatment differences in energy use between treatment and control groups for $\tau = -2, -1, 0$.

[21]Figure A.6 in the Appendix presents analogous plots that include pre-treatment periods $\tau = -6, -5, -4, -3$. These are estimated using households that are eligible for the second treatment wave. Like Figure 3, these plots reveal no statistically or economically-significant pre-treatment differences in levels or trends of energy use between treatment and control.

fixed effects, for treated and control households in the lowest quintile of the pre-treatment use distribution broken down into two groups: low-use households who overestimate their relative energy use, and all other low users (underestimators, households with correct priors, and survey non-respondents). Boomerang effects are present for both groups, but appear to be somewhat stronger among households who overestimate.

Panel (b) of Figure 4 provides an interesting contrast to panel (a). Here, we plot daily average energy use for households in the highest-use quintile, broken down to contrast households who underestimate their relative energy use to all other households in the highest-use quintile. We find, as expected, some initial evidence of high-use customers who underestimate their relative energy use responding more strongly to treatment, and conserving relatively more energy than other high energy users.[22]

5.2 Nonparametric tests for treatment effect heterogeneity

Motivated by Figures 3 and 4, we use regressions to formally test for treatment effect heterogeneity as a function of energy-use levels and (un)informedness over relative energy use. Before employing these linear models, we present two nonparametric tests that lend further support to our investigation into heterogeneous treatment effects.

The first test comes from Crump et al. 2008. They develop a nonparametric test of the null hypothesis that the treatment has zero effect for all subpopulations defined by covariates in the data. The test does not reveal along which dimensions treatment effect heterogeneity exists (e.g., pre-treatment energy use, beliefs, or demographics), just that there exist covariates in the data that generate heterogeneous conditional average treatment effects.[23] In implementing the test, we consider the "all covariates," "top-down," and "bottom-up" specifications from Crump et al. 2008. The covariates considered by the test include: (1) dummies for each quintile of the pre-treatment energy-use distribution; (2) dummies for

[22]For reference, Figure A.6 plots raw average daily use in levels in calendar time without removing fixed effects for the first treatment wave. It similarly depicts heterogeneous treatment effects among low and high use households.

[23]In the interest of space, we refer the reader to Crump et al. 2008 and Ferraro and Miranda 2013 for discussion of the details of the test and its implementation.

whether a household overestimates/is correct about/underestimates their energy-use quintile; (3) a dummy for whether a household responded to the baseline survey; and (4) all the demographic variables in Table 2.

The nonparametric test results for each month between November 2012 and May 2013 are presented in Table 6. The small p-values in each month across all test specifications imply that there exist covariates in the data that generate heterogeneous treatment effects.

For our second nonparametric test, we compute quantile treatment effects.[24] The estimates for the middle post-treatment month, March 2013, are presented in Figure 5; the results are similar for other months. The results are consistent with the graphical evidence from above: households at lower quantiles of the baseline energy use distribution increase their use in response to treatment while those at higher quantiles decrease their use.[25]

[24]Results from Firpo 2007 are relevant for the estimation of quantile treatment effects in the presence of endogenous sample selection. In particular, they show unconditional quantile treatment effects can be recovered if the data are reweighted using IPWs that predict treatment status.

[25]The quantile treatment effects can be interpreted as the distribution of treatment effects if we assume a household's rank in the outcome distribution is the same regardless of whether they are assigned to treatment or control (Bitler et al. 2008). Following Bitler et al. 2008 and Djebbari and Smith 2008, we test this assumption in Table A.3 of the Appendix and find strong evidence against it. We thus caution that Figure 5 should be interpreted solely as the impact of treatment on the distribution and not as the distribution of treatment effects.

5.3 Heterogeneous treatment effects

We now estimate the treatment effect of our information treatment on daily electricity use for household i in month t, q_{it}. We do so with the following regression:

$$
\begin{aligned}
\log(q_{it}) &= \alpha_1 T_{it} + \sum_{j=1}^{3} \alpha_{2,j}(T_{it} \times B_{i,j}) + \sum_{k=1}^{5} \alpha_{3,k}(T_{it} \times Q_{i,k}) + \sum_{d=1}^{D} \alpha_{4,d}(T_{it} \times X_{i,d}) \\
&+ \sum_{j=1}^{3} \delta_{j,1}(t \times 1\{t \geq \text{Oct2012}\} \times B_{i,j}) + \sum_{j=1}^{3} \gamma_{j,1}(t \times 1\{t \geq \text{Mar2013}\} \times B_{i,j}) \\
&+ \sum_{k=1}^{5} \delta_{k,2}(t \times 1\{t \geq \text{Oct2012}\} \times Q_{i,k}) + \sum_{k=1}^{5} \gamma_{k,2}(t \times 1\{t \geq \text{Mar2013}\} \times Q_{i,k}) \\
&+ \sum_{d=1}^{D} \delta_{d,3}(t \times 1\{t \geq \text{Oct2012}\} \times X_{i,d}) + \sum_{d=1}^{D} \gamma_{d,3}(t \times 1\{t \geq \text{Mar2013}\} \times X_{i,d}) \\
&+ \mu_i + \tau_t + \epsilon_{it}. \qquad (2)
\end{aligned}
$$

The treatment variable, T_{it}, equals one if household i is sent the information treatment in month t and zero otherwise. The α_1 coefficient corresponds to an Intent-to-Treat (ITT) effect since not all treatment households comply and engage with the emails and web portal.

We allow for heterogeneous treatment effects with the next three sets of terms in equation (2). The beliefs dummies $B_{i,1}, B_{i,2}, B_{i,3}$ equal one if household i overestimates, is correct about, or underestimates their relative electricity use, respectively. The omitted category is survey non-respondents. The dummy $Q_{i,k}$ equals one if household i's pre-treatment electricity use is in quintile k. The vector $X_{i,d}$ corresponds to the demographic variables in Table 2. For demographic d, we include a dummy variable in $X_{i,d}$ that equals one if household i lives in a census block whose value for demographic d is above the sample median across all census blocks.[26] Our empirical specification thus allows us to identify how household's

[26]For instance, if household i's census block has average household income that is above the median household income, then the dummy equals one, indicating that the household lives in a "high" income area. The dummies for age, employment, education, and Green Party local-vote share are similarly defined. This specification of demographic variables follows Ferraro and Miranda 2013. In practice we codify demographics in this way, rather than simply including the raw demographics in $X_{i,d}$, as it yields a better model fit to the data. Our main findings are unchanged if we instead include the raw

pre-treatment beliefs affect their response to treatment ($\alpha_{2,j}$), while allowing for treatment responses to simultaneously depend on a household's level of electricity use ($\alpha_{3,k}$) and demographics ($\alpha_{4,d}$).

Mean reversion and seasonality in energy use as a function of beliefs, pre-treatment energy-use levels, or demographics are potentially major confounds to the identification of $\alpha_{2,j}, \alpha_{3,k}, \alpha_{4,d}$. To account for these factors, we allow for differential post-treatment trends in energy use as a function of pre-treatment beliefs $B_{i,j}$, baseline-use quintile $Q_{i,k}$, and demographics $X_{i,d}$. We do so by including post-treatment indicators with these respective variables in the second, third, and fourth lines of equation (2). Importantly, we allow for differential trends following both the October 2012 and March 2013 treatment waves. The $\delta_{k,1}, \delta_{k,2}, \delta_{k,3}$ coefficients capture differential energy use by observable characteristics between October and February, while the $\gamma_{k,1}, \gamma_{k,2}, \gamma_{k,3}$ coefficients correspond to energy use in March through May.

Finally, we account for household fixed effects and month fixed effects in all specifications.

Results

Table 7 presents our findings. Throughout, we report coefficient estimates from equation (2) multiplied by 100 so that the values in the table directly correspond to percentage changes in energy use. The ITT estimates in columns (1)–(5) are largely consistent with the patterns in Figures 3 and 4. The pooled ITT estimate yields a null result from our information treatment that masks treatment heterogeneity. Our preferred specification in column (5) reveals that households who overestimate their relative energy use prior to the experiment increase their energy use by 6.3% once they are offered treatment. In contrast, the conservation effect following treatment among high users who underestimate relative energy use that we observe in panel (b) of Figure 4 does not survive once we control for other moderators of treatment and heterogeneous post-treatment trends in energy use.

We also find important differences in treatment effects among high and low energy users. Households in the lowest quintile of the baseline energy-use distribution have an ITT of 11.7% while the ITT for households in the highest quintile is -11%.[27]

demographics in $X_{i,d}$.

[27]For the interested reader, we report Local Average Treatment Effects in Table A.4 in the Appendix.

Robustness checks

Recall that the results shown in Table 7 include IPWs to account for potential attrition bias. Table A.5 in the Appendix presents results without the attrition correction. The results are near identical. Our preferred specification in column (5) reveals that households who overestimate their relative energy use prior to the experiment increase their energy use by 6.48% (instead of 6.30%), and that low users increase their consumption by 11.09% (instead of 11.71%).

One remaining concern is there might be time-varying unobservables that are correlated with attrition. IPWs, which are based on observables, would not account for these. To explore this concern we have conducted an analysis in the vein of Altonji, Elder and Taber 2005 (available upon request). We find that any omitted variable bias, potentially related to attrition, would have to be implausibly large relative to the impact of the control variables included in Eq. (2) on our coefficients of interest in order for our main results to be invalid.

5.4 Time-varying treatment effects

Does treatment heterogeneity among households with different pre-treatment beliefs and energy-use persist over time? Figures 6 and 7 address this question. The figures plot the monthly treatment effect (ITT) by the number of months before or since a household was first emailed about the smart-meter portal in event time. To construct these figures, we include a full set of interactions between three months-before-treated and seven months-since-treated dummies and all the terms that include T_{it} in equation (2), and estimate the time-varying treatment effects.[28] Figures 6 and 7 thus depict the time-varying $\alpha_{2,j}$ and $\alpha_{3,k}$

In estimating the LATE, we codify household i as complying with treatment in month t if it opens a Billcap email and views the summary of their recent energy use and costs at least once in month t. We find a 7% incremental increase in energy use from engaging with smart-meter data among households who overestimate their relative use. Low energy users in the first quintile of the use distribution increase their use by 16.5% from engaging with their smart-meter data. In contrast, high users in the fifth quintile decrease their use by 16%.

[28] As in Figure 3, we estimate time-varying treatment effects for periods $\tau = -2, -1, 0$ before treatment using all households for July, August and September 2012, and all households for December 2012,

estimates from equation (2).

Figure 6 again emphasizes the importance of pre-treatment beliefs in influencing the impact of our information treatment. Only households who overestimate their relative energy use (panel (a)) exhibit a differential increase in energy use in response to treatment. However, the figure further reveals this effect is temporary and becomes statistically insignificant six months after treatment. One interpretation of this result is that becoming informed about relative use generates a short-lived salience effect among overestimators. We do not find persistent changes in behavior from correcting errors-in-beliefs about relative use that would, for example, be predicted by models of social learning.

Figure 7 shows, in contrast, that the treatment effect heterogeneity as a function of baseline energy use is persistent. This is highlighted in panels (a) and (d) for households in the lowest and highest quintiles of the energy-use distribution: after controlling for prior beliefs, low-use households exhibit a persistent 10% increase in energy use due to treatment while high-use households exhibit a similar-magnitude, persistent decrease in energy use.

What generates the patterns in Figure 7 if they are not being driven by households responding to new information from peer comparisons? The other important aspect of our treatment – higher-frequency and more granular information – may help explain differential treatment effects among low and high energy users. Relative to control households that receive quarterly bills, treatment households receive detailed, daily or weekly smart-meter information at least every two weeks. Among high energy users, salience may play a role by making costly electricity behavior more often front-of-mind.[29] Further, the energy-saving tips that accompany the information treatment could drive some of the conservation effect from treatment among the high users in panel (b) of Figure 7.

Individual learning about how daily electricity use translates into bills may also play a role. Although households on average underestimate the electricity use of large appliances

---

January and February 2013 who are not treated in the first treatment wave and who do not attrit between December 2012 and March 2013

[29]Another psychological factor that might influence our treatment effects is activation of self identities. Confirmation that one is a high or low user through our information treatment could lead such households to increasingly act like high or low users.

(Attari et al. 2010), high users may also have previously overestimated the cost of behavioral changes that reduce energy use and electricity bills. Similarly, the sustained boomerang effect in panel (a) of Figure 7 could be explained by low-use households learning that their daily energy-use behaviors are less costly financially than previously believed. For instance, upon learning from their smart-meter data that ensuring all unused lights in the home are always off saves them less than the cost of a postage stamp each week, low users may scale back their conservation efforts, causing a boomerang effect.

Of course, there are other possible interpretations of the boomerang effect in panel (a). For instance, according to Brehm's theory of psychological reactance, a post-treatment increase in energy use could be due to a psychological opposition effect driven by resentment at being told what to do. Although we can not rule out that a knee-jerk oppositional reaction is driving the increase among low users, the persistent nature of the boomerang effect suggests that households are reoptimizing in the presence of new information on energy use. The distinction is important: reoptimization suggests our information treatment is welfare-improving, whereas an opposition reaction represents a psychological cost. To the extent that the boomerang effects come from social or individual learning, information treatments may be welfare-improving even if they result in higher energy use.

6 Conclusion

In this paper, we use a field experiment in retail electricity to study the impacts of an information program with peer comparisons. The key innovation of our experiment is that we elicit households' beliefs about relative levels of pro-social behavior with a pre-treatment survey, and then inform these beliefs through an information treatment. Our analysis delivers three main results. First, before seeing their smart-meter data, households do not systematically underestimate their relative electricity use. Rather, most households believe they are "average" energy users, with high and low users underestimating and overestimating their relative levels of energy use at symmetric rates.

Despite the symmetry in errors-in-beliefs, treatment effects from our information program vary asymmetrically. We find that errors-in-beliefs regarding relative use can generate unintended consequences: namely, increases in use by underestimators, which is a type of boomerang effect. Correcting errors-in-beliefs about relative use does not lead to increased energy conservation. The implication for policy aimed at conservation is that peer-comparison information should be targeted toward those who are initially failing to meet the social norm, more specifically those who are more likely to overestimate and not underestimate their prosocial behavior.

Finally, we find a second and larger boomerang effect from our information treatment that is unrelated to errors-in-beliefs. Controlling for beliefs, low-use households persistently increase energy use in response to treatment. At the same time, high-use treatment households persistently decrease energy use. While we are unable to confirm the mechanisms driving these heterogeneous effects, their persistence points to individual learning and reoptimization. Regardless of the mechanism, if the goal of information programs is to promote prosocial behavior, our results underscore the increasing importance of targeting information as future programs exploit high-frequency, personalized data and information technology.

References

Allcott, Hunt. "Social norms and energy conservation." *Journal of Public Economics* 95, 9: (2011) 1082–1095.

———. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 130, 3: (2015) 1117–1165.

Allcott, Hunt, and Judd B. Kessler. "The welfare effects of nudges: A case study of energy use social comparisons.", 2015. National Bureau of Economic Research WP 21671.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 13, 1: (2005) 151–184.

Ashraf, Nava, Oriana Bandiera, and Scott S Lee. "Awards unbundled: Evidence from a natural field experiment." *Journal of Economic Behavior & Organization* 100 IS -: (2014) 44–63.

Attari, Shahzeen Z, Michael L DeKay, Cliff I Davidson, and Wändi Bruine de Bruin. "Public perceptions of energy consumption and savings." *Proceedings of the National Academy of Sciences* 107, 37: (2010) 16,054–16,059.

Ayres, Ian, Sophie Raseman, and Alice Shih. "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage." *Journal of law, economics and organization* 0: (2012) 1–31.

Becker, Gary. "A Theory on the Allocation of Time." *Economic Journal* 75: (1965) 493–517.

Bernedo, María, Paul J Ferraro, and Michael Price. "The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation." *Journal of Consumer Policy* 37, 3: (2014) 437–452.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. "The effect of providing peer information on retirement savings decisions." *Journal of Finance* 70, 3: (2015) 864–882.

Bitler, Marianne, Jonah Gebalch, and Hilliary Hoynes. “Distributional Impacts of the Self-Sufficiency Project.” *Journal of Public Economics* 92, 3-4: (2008) 748–765.

Bitler, Marianne P., Jonah B Gelbach, and Hilary W. Hoynes. “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments.” *American Economic Review* 96, 4: (2006) 988–1012.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen. “Calorie posting in chain restaurants.” *American Economic Journal: Economic Policy* 3: (2011) 91–128.

Brent, Daniel A, Joseph H Cook, and Skylar Olsen. “Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials.” *Journal of the Association of Environmental and Resource Economists* 2, 4: (2015) 597–627.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini. “Overconfidence and Social Signalling.” *Review of Economic Studies* 80: (2013) 949–983.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation.” *Econometrica: Journal of the Econometric Society* 81, 3: (2013) 855–882.

Chetty, Raj, Adam Looney, and Kory Kroft. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99, 4: (2009) 1145–1177.

Costa, Dora L., and Matthew E. Kahn. “Energy conservation “nudges” and environmentalist ideology: Evidence from a randomized residential electricity field experiment.” *Journal of the European Economic Association* 11, 3: (2013) 680–702.

Crump, Richard K., Joseph V. Hotz, Guido W. Imbens, and Oscar A. Mitnik. “Overconfidence and Social Signalling.” *Review of Economics and Statistics* 90: (2008) 389–405.

Davis, Lucas W, and Erich Muehlegger. “Do Americans consume too little natural gas? An empirical test of marginal cost pricing.” *The RAND Journal of Economics* 41, 4: (2010) 791–810.

DellaVigna, Stefano, and Ulrike Malmendier. “Paying Not to Go to the Gym.” *American Economic Review* 96: (2006) 694–719.

Delmas, Magali A, Miriam Fischlein, and Omar I Asensio. “Information Strategies and Energy Conservation Behavior: A Meta-analysis of Experimental Studies from 1975-2012.” *Energy Policy* 61: (2013) 729–739.

Djebbari, Habiba, and Jeffrey Smith. “Heterogeneous Impacts of PROGRESA.” *Journal of Econometrics* 145: (2008) 64–80.

Duflo, Esther, and Emmanuel Saez. “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment.” *Quarterly Journal of Economics* 118, 3: (2003) 815–842.

Dunning, David, Chip Heath, and Jerry M Suls. “Flawed Self-Assessment: Implications for Health, Education, and the Workplace.” *Psychological Science in the Public Interest* 5, 3: (2004) 69–106.

Dupas, Pascaline. “Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya.” *American Economic Journal: Applied Economics* 3: (2011) 1–34.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. “Testing enforcement strategies in the field: Threat, moral appeal and social information.” *Journal of the European Economic Association* 11, 3: (2013) 634–660.

Ferraro, Paul J., and Juan J. Miranda. “Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment.” *Resource and Energy Economics* 35, 3: (2013) 356–379.

Ferraro, Paul J., and Michael K. Price. “Using non-pecuniary strategies to influence behavior: Evidence from a large scale field experiment.” *Review of Economics and Statistics* 95, 1: (2013) 64–73.

Firpo, Sergio. “Efficient Semiparametric Estimation of Quantile Treatment Effects.” *Econometrica* 75, 1: (2007) 259–276.

Frey, Bruno S., and Stephan Meier. "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment." *American Economic Review* 94, 5: (2004) pp. 1717–1722.

Gerber, Alan S., and Todd Rogers. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You." *Journal of Politics* 71, 1: (2009) 178–191.

Gilovich, Thomas. *How we know what isn't so.* Simon and Schuster, 2008.

Goette, Lorenz, Verena Tiefenbeck, Kathrin Degen, Vojkan Tasic, Rafael Lalive, Elgar Fleisch, and Thorsten Staake. "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation." *Management Science* Forthcoming.

Heckman, James J. "Sample Selection Bias as a Specification Error." *Econometrica* 47, 1: (1979) 153–161.

Jalan, Jyotsna, and E Somanathan. "The importance of being informed: Experimental evidence on demand for environmental quality." *Journal of Development Economics* 87, 1: (2008) 14–28.

Jensen, Robert. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125, 2: (2010) 515–548.

Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel. "Salience and Taxation: Theory and Evidence." *Quarterly Journal of Economics* 127, 1: (2012) 199–235.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126, 1: (2011) 145–205.

Kruger, Justin. "Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments." *Journal of personality and social psychology* 77, 2: (1999) 221.

Levitt, Steven D, and John A List. "What Do Field Experiments Measuring Social Preferences Reveal About The Real World?" *Journal of Economic Perspectives* 21, 2: (2007) 153–174.

Luby, Stephen P, Mubina Agboatwalla, Daniel R Feikin, John Painter, Ward Billhimer, Arshad Altaf, and Robert M Hoekstra. "Effect of Handwashing on Child Health: A Randomised Controlled Trial." *Lancet* 366, 9481: (2005) 225–233.

Mountain, Bruce. "Electricity Prices in Australia: An International Comparison." Technical report, Carbon and Energy Markets CME, 2012.

Opower. "Think you're better than others at saving energy? 3 charts suggest you may need a reality check.", 2014. https://blog.opower.com/2014/04/think-youre-better-than-others-at-saving-energy-these-3-charts-suggest-you-may-need-a-reality-check/.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. "The constructive, destructive, and reconstructive power of social norms." *Psychological Science* 18, 5: (2007) 429–434.

Svenson, Ola. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica* 47, 2: (1981) 143–148.

Wichman, Casey. "Information Provision and Consumer Behavior: A Natural Experiment in Billing Frequency." *Journal of Public Economics* 152, 1: (2017) 13–33.

Williams, Elanor F, and Thomas Gilovich. "Do people really believe they are above average?" *Journal of Experimental Social Psychology* 44, 4: (2008) 1121–1128.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Parienté, Clair Null, Edward Miguel, Michael Kremer, Dean S Karlan, Richard Hornbeck, and Xavier Giné. "Being surveyed can change later behavior and related parameter estimates." *Proceedings of the National Academy of Sciences* 108, 5: (2011) 1821–1826.

A Tables

Table 1: Average Daily Pre-Treatment Energy Use (kWh) by Home Energy Report Treatment Wave

	Treatment waves			Treatment waves		
	Oct 2012	Mar 2013 & Jun 2013	Diff.	Mar 2013	Jun 2013	Diff.
Jul 2012	14.16	13.88	0.29			
			(0.33)			
Aug 2012	13.63	12.94	0.70			
			(0.45)			
Sep 2012	11.52	10.95	0.57			
			(0.34)			
Oct 2012				9.80	9.94	-0.14
						(0.26)
Nov 2012				8.96	9.09	-0.13
						(0.23)
Dec 2012				8.79	9.05	-0.26
						(0.24)
Jan 2013				9.27	9.48	-0.21
						(0.27)
Feb 2013				9.92	9.62	0.30
						(0.28)

Notes: Means and difference in means in average daily electricity use (in kWh) between treatment waves are reported. Population of residential households without solar panels whose energy use does not exceed 50 kWh/day with smart-meter readings dating back to at least October 1, 2012. Standard errors of differences in means are in parentheses and are clustered at the SA1 level. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table 2: Household Characteristics and Census Block-level Demographics by Home Energy Report Treatment Wave

	Treatment wave			Difference in means		
	Oct 2012	Mar 2013	Jun 2013	Oct–Mar	Oct–Jun	Mar–Jun
(a) Survey data						
Has air conditioning	0.71	0.62	0.61	0.08	0.10	0.01
				(0.06)	(0.10)	(0.10)
Has gas appliances	0.70	0.60	0.57	0.10*	0.13	0.03
				(0.06)	(0.10)	(0.10)
Has swimming pool	0.02	0.02	0.00	-0.00	0.02	0.02*
				(0.02)	(0.01)	(0.01)
Has freestanding house	0.49	0.34	0.25	0.15***	0.24***	0.09
				(0.06)	(0.09)	(0.09)
Number of residents	2.23	2.18	2.07	0.05	0.16	0.11
				(0.13)	(0.21)	(0.20)
Number of bedrooms	2.41	2.20	2.04	0.21*	0.38**	0.17
				(0.11)	(0.18)	(0.18)
(b) Census data						
Median weekly household income	1432.53	1423.80	1335.93	8.73	96.61***	87.87***
				(28.20)	(32.58)	(29.17)
Average age	37.40	37.13	36.48	0.27	0.92***	0.66**
				(0.32)	(0.36)	(0.33)
Full-time employment rate	42.42	43.02	40.97	-0.60	1.45*	2.05***
				(0.67)	(0.81)	(0.66)
Proportion of home renters	37.66	39.37	40.76	-1.71*	-3.10*	-1.39
				(1.03)	(1.24)	(1.05)
Has above median vote share for Green Party	0.53	0.55	0.52	-0.02	0.01	0.03
				(0.03)	(0.03)	(0.02)
Number of households	640	853	930			

Notes: Population of residential households without solar panels whose energy use does not exceed 50 kWh/day with smart-meter readings dating back to at least October 1, 2012. All variables that start with "Has" are dummies that equal one if the answer is yes and zero otherwise. Means and differences in means between treatment waves reported. Standard errors of differences in means are in parentheses and are clustered at the SA1 level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3: Summary Statistics by Survey Response and Attrition Status

	Survey respondents	Non-respondents	Diff.	Attritors	Non-attritors	Diff.
(a) Average pre-treatment daily energy use (kWh)						
Jul 2012	13.83	13.99	-0.16	14.05	13.87	0.18
			(0.44)			(0.31)
Aug 2012	13.00	13.16	-0.16	13.09	13.19	-0.10
			(0.41)			(0.29)
Sep 2012	11.13	11.12	0.01	11.11	11.13	-0.02
			(0.32)			(0.24)
(b) Census data						
Median household income	1444.20	1384.86	59.34*	1452.57	1332.11	120.46***
			(34.96)			(28.65)
Average age	37.79	36.83	0.96*	37.21	36.69	0.52
			(0.40)			(0.39)
Full-time employment rate	42.73	41.98	0.76	43.63	40.51	3.12***
			(0.80)			(0.85)
Proportion of households renting	37.75	39.70	-1.95	39.33	39.57	-0.24
			(1.21)			(1.27)
Above median vote share for Green Party	0.56	0.53	0.02	0.57	0.49	0.08***
			(0.03)			(0.02)
Number of households	311	2112		1028	1395	

Notes: All variables that start with "Has" are dummies that equal one if the answer is yes and zero otherwise. Means and difference in means between treatment waves reported. Standard errors of differences in means are in parentheses and are clustered at the SA1 level. $^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$.

Table 4: Joint Distribution of Prior Beliefs and Actual Relative Energy Use

(a) All survey respondents (1188 observations)

		Pre-treatment belief					
		1-20%	20-40%	40-60%	60-80%	80-100%	Total
Actual energy use quintile	1-20%	**1.86**	6.84	10.21	1.01	0.17	20.08
	20-40%	0.68	**4.89**	11.98	2.11	0.25	19.92
	40-60%	0.84	3.12	**12.74**	2.78	0.76	20.25
	60-80%	0.68	3.46	11.56	**3.71**	0.93	20.34
	80-100%	0.34	2.28	11.56	3.71	**1.52**	19.41
	Total	4.39	20.59	58.06	13.33	3.63	100.00

(b) Experimental sample (311 observations)

		Pre-treatment belief					
		1-20%	20-40%	40-60%	60-80%	80-100%	Total
Actual energy use quintile	1-20%	**2.26**	6.77	9.35	0.97	0.32	19.68
	20-40%	0.97	**6.45**	10.97	1.61	0.32	20.32
	40-60%	0.97	3.55	**12.26**	2.58	0.32	19.68
	60-80%	0.65	4.84	12.90	**2.58**	0.65	21.61
	80-100%	0.00	3.23	12.90	1.61	**0.97**	18.71
	Total	4.84	24.84	58.39	9.35	2.58	100.00

Notes: The prior beliefs are reported quintiles conditional on similar sized homes from our pre-treatment baseline survey. The actual use quintiles correspond to pre-treatment energy use conditional on the same number of bedrooms. The experimental sample is restricted to households with smart-meter reads as of October 1, 2012.

Table 5: Bayesian Model Allocation Functions

(a) Empirical allocation function $q_k(s_j)$

Actual Energy Use Quintile	1-20%	20-40%	40-60%	60-80%	80-100%
1-20%	**9.24**	34.03	50.84	5.04	0.84
20-40%	3.36	**24.37**	59.66	10.50	1.26
40-60%	4.20	15.55	**63.45**	13.87	3.78
60-80%	3.36	17.23	57.56	**18.49**	4.62
80-100%	1.68	11.34	57.56	18.49	**7.56**

(b) Constrained ML allocation function estimates $q_k^{ML}(s_j)$

Actual Energy Use Quintile	1-20%	20-40%	40-60%	60-80%	80-100%
1-20%	**9.96**	29.81	53.89	5.43	0.91
20-40%	3.17	**29.81**	56.71	9.12	1.19
40-60%	4.13	14.88	**63.22**	14.05	3.72
60-80%	3.28	17.22	56.17	**18.82**	4.51
80-100%	1.75	12.22	59.36	18.82	**7.86**

Notes: Results are based on the full sample of 1188 survey respondents. The empirical allocation in panel (a) reports for each pre-treatment electricity-use quintile k the fraction of households that report, pre-treatment, they are in electricity-use quintile j. The constrained-maximum likelihood allocation function estimates in panel (b) report the solution to the problem described in equation (1) in the text. These estimates simultaneously best-fit the data while satisfying restrictions from the model of Bayesian updating from Burks et al. 2013; see the text for details.

Table 6: Tests for Zero Conditional Average Treatment Effect

Period	Covariates selection	*chi*-sq	(dof)	p-val	normal	p-val
Nov 2012	AC	45.92	13	0.00	6.46	0.00
	TD	36.50	8	0.00	7.12	0.00
	BU	35.88	9	0.00	6.34	0.00
Dec 2012	AC	14.29	9	0.11	1.25	0.11
	TD	13.55	7	0.06	1.75	0.04
	BU	35.24	9	0.00	6.19	0.00
Jan 2013	AC	19.53	9	0.02	2.48	0.01
	TD	14.21	5	0.01	2.91	0.00
	BU	26.93	10	0.00	3.79	0.00
Feb 2013	AC	14.19	8	0.08	1.55	0.06
	TD	10.34	6	0.11	1.25	0.11
	BU	32.35	10	0.00	5.00	0.00
Mar 2013	AC	15.39	8	0.05	1.85	0.03
	TD	11.04	6	0.09	1.45	0.07
	BU	14.60	12	0.26	0.53	0.30
Apr 2013	AC	11.37	7	0.12	1.17	0.12
	TD	4.73	4	0.32	0.26	0.40
	BU	12.34	7	0.09	1.43	0.08
May 2013	AC	29.36	11	0.00	3.91	0.00
	TD	13.92	7	0.05	1.85	0.03
	BU	9.00	5	0.11	1.26	0.10

Notes: AC, TD, BU respectively denote "all covariates," "top down," and "bottom up" covariates selection for implementing the nonparametric test of zero conditional average treatment effects from Crump et al. 2008. The "*chi*-sq" and "normal" columns correspond to the chi-square and normal versions of these tests. See the text of Crump et al. 2008 for details on these tests and their implementation. For Nov 2012 – Feb 2013, the test is implemented using all households who were with the retailer before October 2012. For Mar 2013 – May 2013, the test is implemented using households who were with Click Energy before October 2012, but who were not treated in the October 2012 treatment wave. The *chi*-sq column is equal to $\sqrt{2K}$ times the normal column plus K, where K is the degrees of freedom.

Table 7: Treatment Effects of Home Energy Reports on Energy Use by Informedness, Baseline Energy Use and Demographics

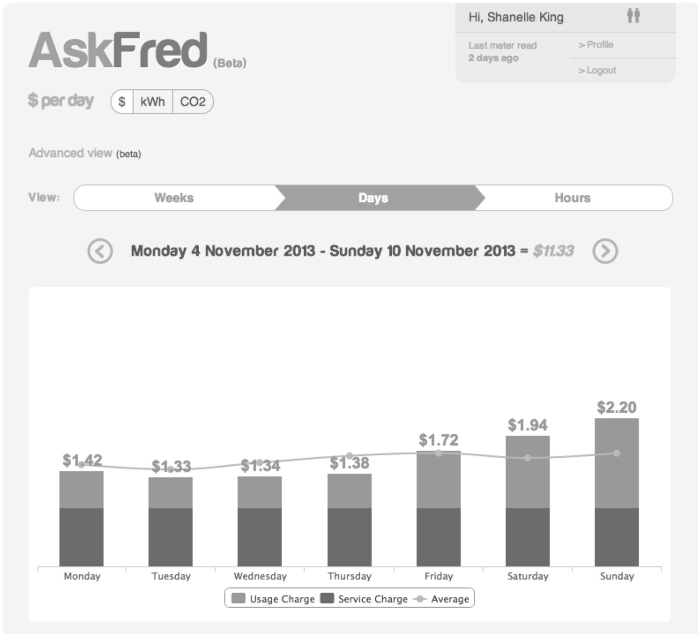
	Dependent variable: log (mean daily kWh)				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.31	0.27	0.96	5.51***	5.23***
	(0.90)	(0.94)	(1.28)	(1.47)	(1.64)
(Underestimate relative energy use)		-7.46*	-0.37	-5.64	0.20
		(3.96)	(4.09)	(3.93)	(4.01)
(Correct about relative energy use)		-1.10	-0.28	0.81	0.72
		(5.27)	(5.09)	(4.97)	(4.85)
(Overestimate relative energy use)		11.57***	5.97**	12.87***	6.30**
		(2.69)	(2.74)	(2.72)	(2.74)
(Low energy use: 1^{st} quintile)			10.67***		11.71***
			(1.95)		(2.09)
(Below avg energy use: 2^{nd} quintile)			1.87		3.10
			(2.07)		(2.16)
(Above avg energy use: 4^{th} quintile)			-6.73***		-5.74***
			(2.24)		(2.39)
(High energy use: 5^{th} quintile)			-11.94***		-10.98***
			(2.43)		(2.62)
(High income)				-2.71	-2.82
				(1.96)	(1.89)
(High age)				-2.71*	-1.77
				(1.58)	(1.67)
(High employment rate)				-3.59*	-1.44
				(2.12)	(2.05)
(High home rental rate)				-4.78***	-5.78***
				(1.84)	(1.79)
(High Green Party support)				11.71	11.28
				(8.67)	(7.98)
Heterogeneous post-treatment trends in electricity use					
Informedness	N	Y	Y	Y	Y
Pre-treatment use	N	N	Y	N	Y
Census (SA1) characteristics	N	N	N	Y	Y
R^2	0.22	0.22	0.24	0.23	0.24
Observations	20129	20129	20129	20129	20129
Treatment households	1531	1531	1531	1531	1531
Control households	892	892	892	892	892

Notes: All specifications include household and month-of-year fixed effects and are weighted by attrition-correction IPWs. Standard errors are clustered at the household level using a pairs cluster bootstrap that accounts for estimation error in implementing IPWs. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. The treatment variable in columns (1)–(5) is a dummy variable equalling one if a household is offered access to the smart-meter web portal. All variables in brackets correspond to interactions of that variable with the treatment dummy variable. Omitted category is survey non-respondents. The dummy variables for the energy-use quintiles correspond to baseline (pre-treatment) energy-use quintiles. High Income is a dummy that equals one if a household lives in an SA1 location whose average income is above median SA1 average income for all SA1s in the sample. The other "High" demographic variables are similarly defined; see the text for details.

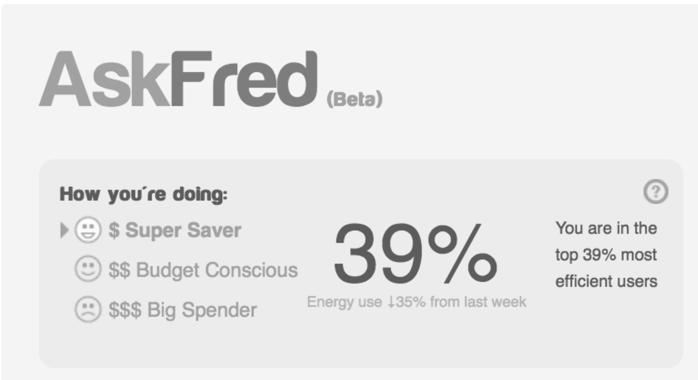
Figure 1: Treatment: Online Home Energy Reports

(a) Homepage

(b) Daily energy use and cost



(c) Peer comparisons



(d) Energy-saving tips

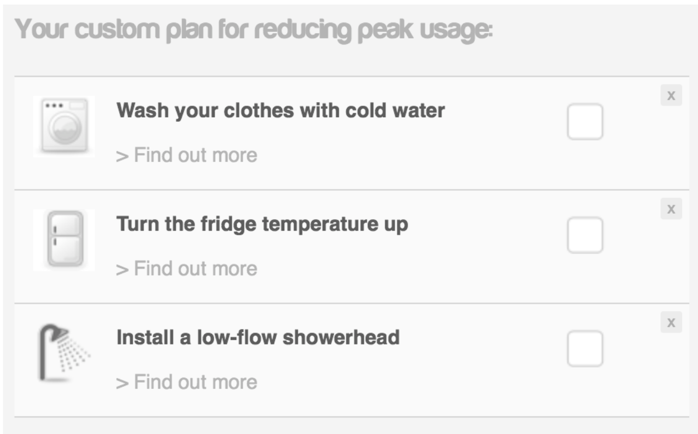
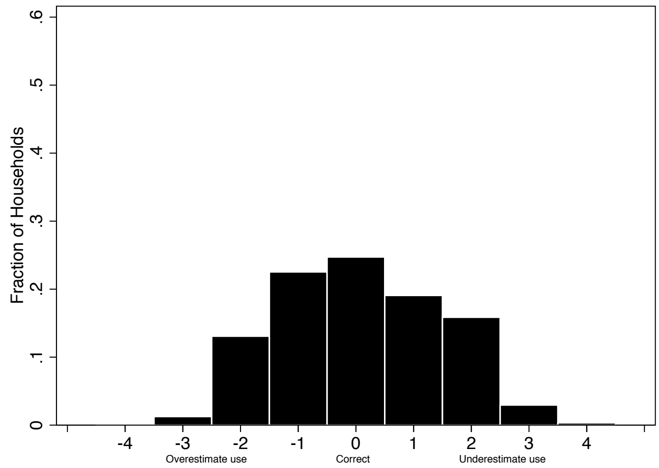
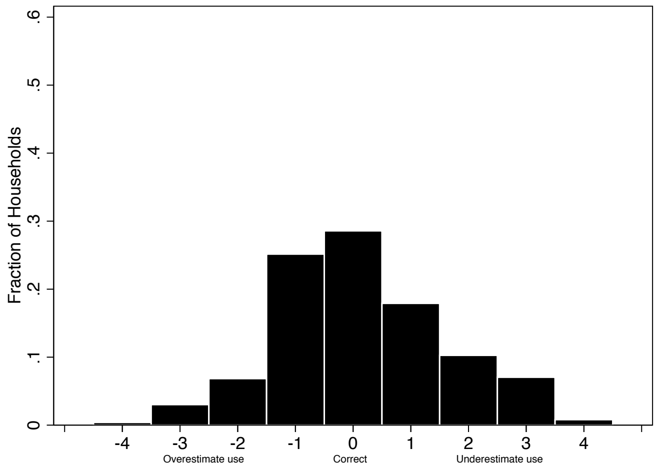


Figure 2: Distribution of Errors-in-beliefs About Relative Energy Use

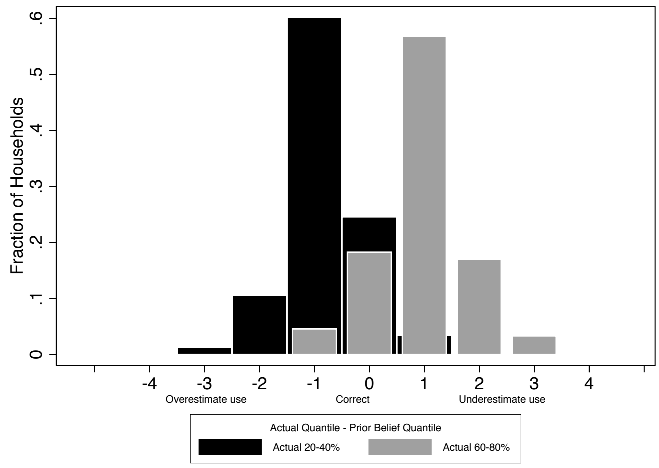
(a) All households



(b) Households without “I’m average” beliefs



(c) Households in 20–40% and 60–80% use quintiles



(d) Households in 1–20% and 80–100% use quintiles

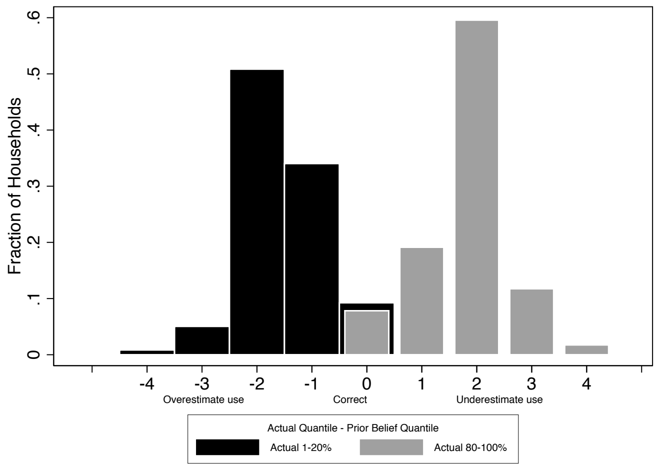
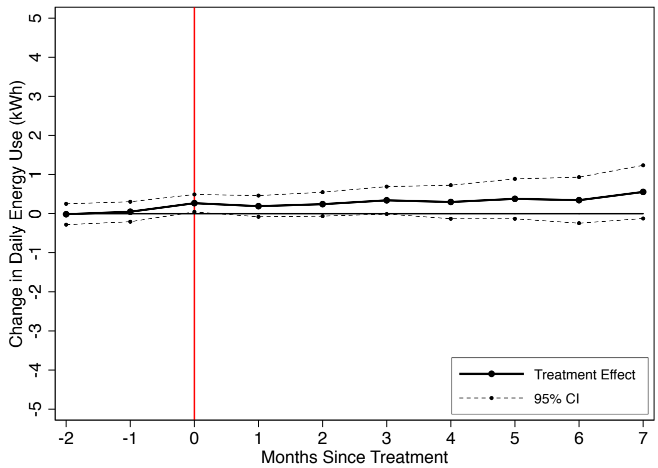
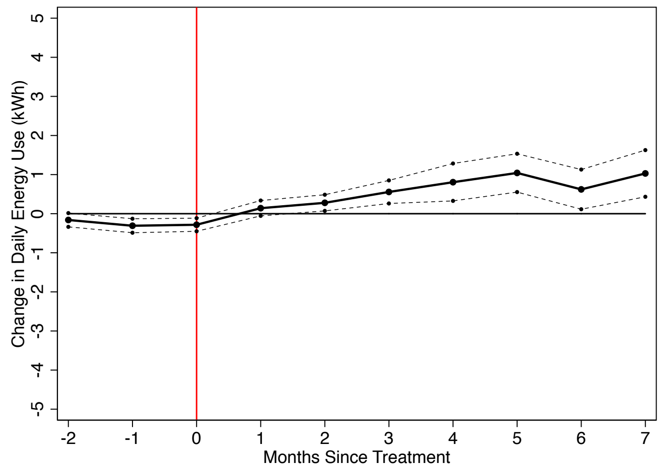


Figure 3: Difference in Average Daily Energy Use between Treatment and Control Within Energy Use Quintile

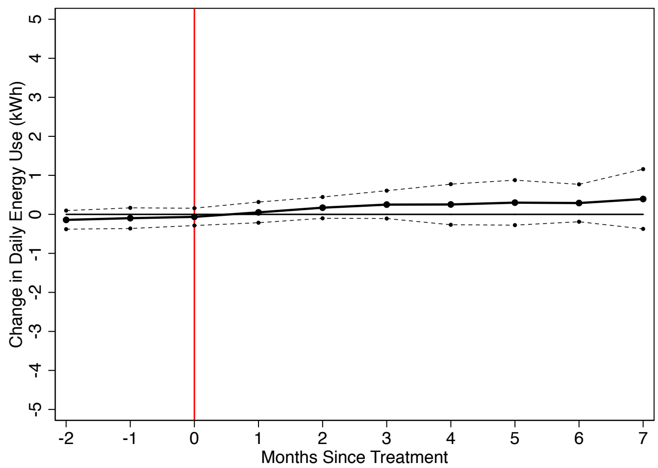
(a) Entire sample



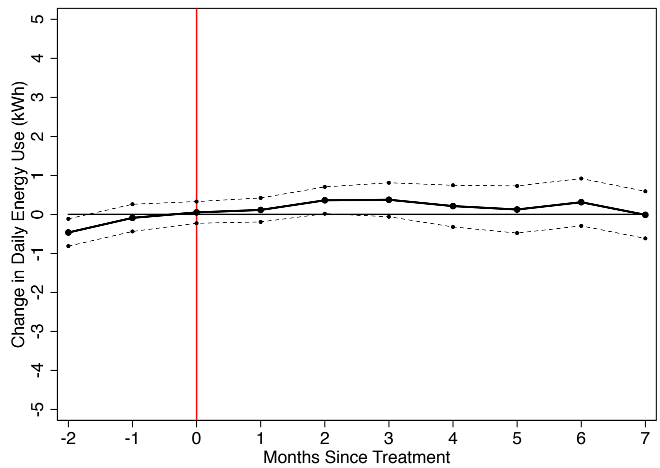
(b) Low users: 0–20% quintile



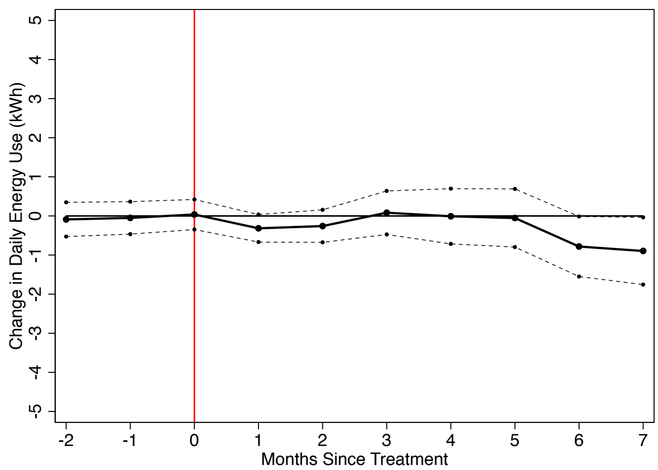
(c) Below average users: 20–40% quintile



(d) Average users: 40–60% quintile



(e) Above average users: 60–80% quintile



(f) High users: 80–100% quintile

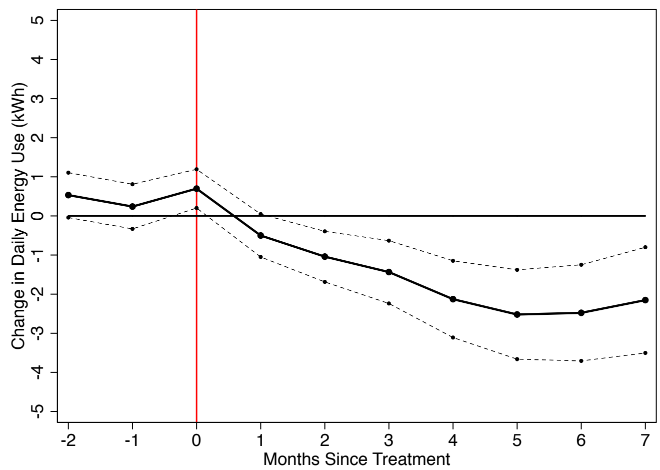
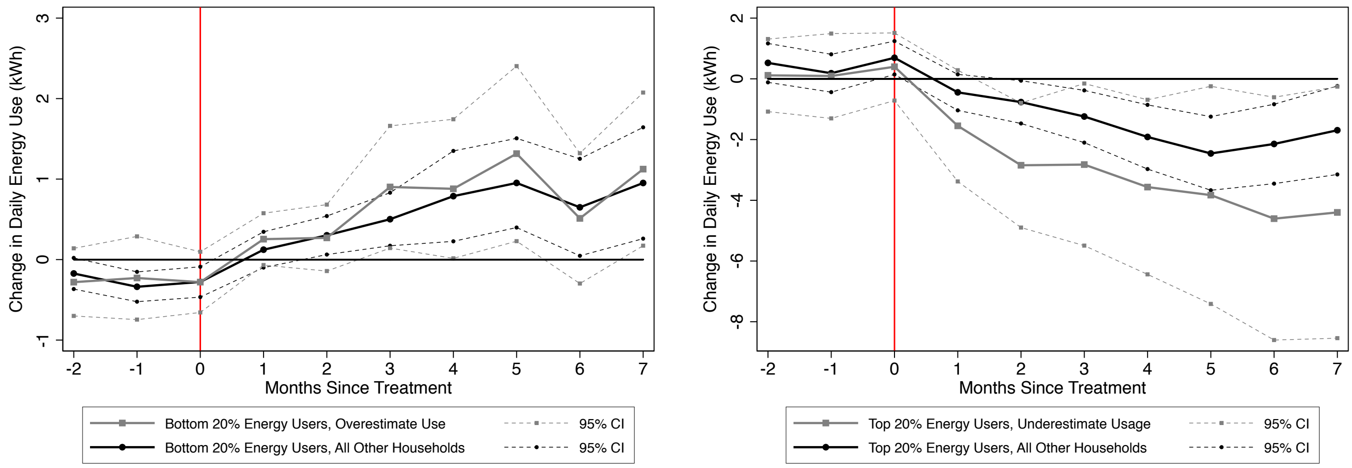


Figure 4: Difference in Average Daily Energy Use Between Treatment and Control Within Use Level and Informedness Sub-group

(a) Low users: 0–20% quintile

(b) High users: 80–100% quintile



Notes: In panel (a) “all other households” is composed of households in the bottom energy use quintile that are correct about use, and those who did not answer the survey. In panel (b) “all other households” is composed of households in the top energy use quintile that are correct about use, and those who did not answer the survey.

Figure 5: Quantile Treatment Effects

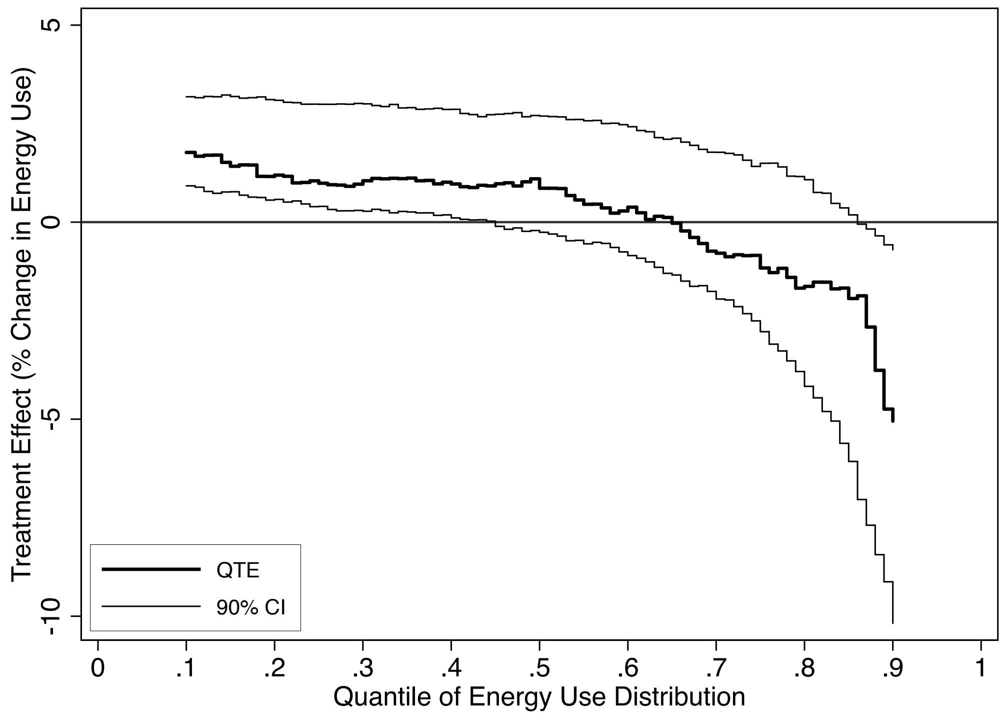
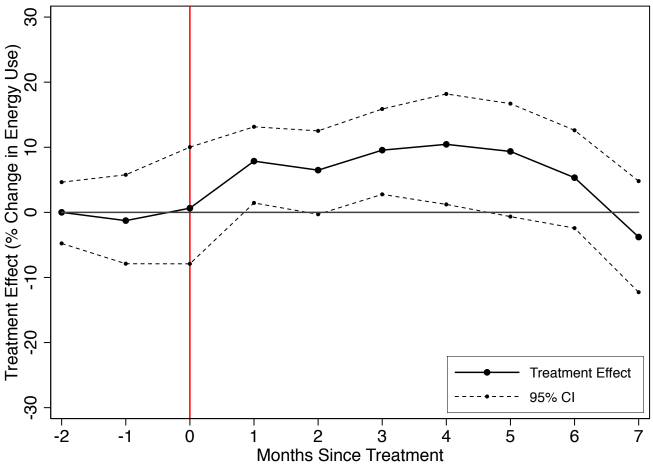
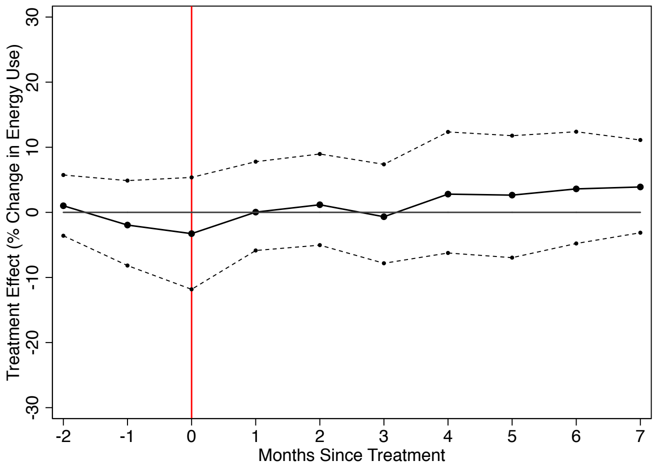


Figure 6: Time Varying Treatment Effects by Informedness

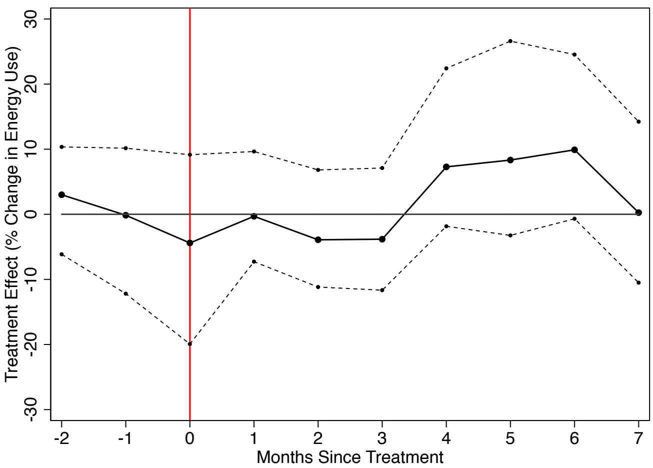
(a) Overestimate relative energy use



(b) Underestimate relative energy use



(c) Correct about relative energy use



(d) Non-respondents to baseline survey

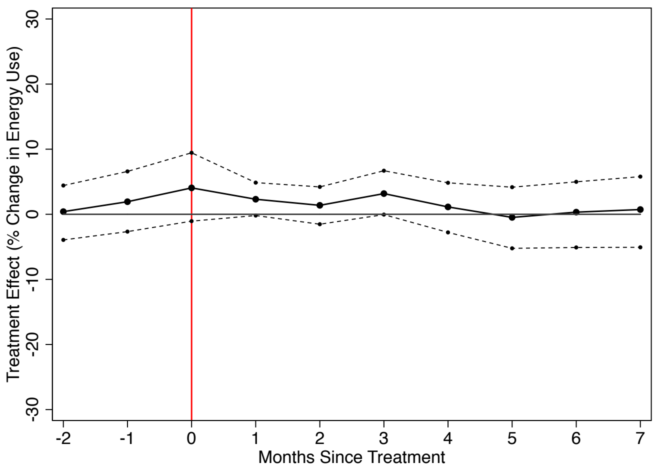
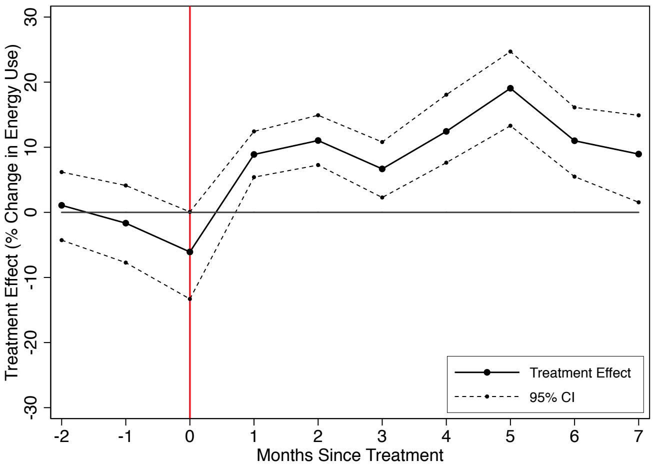
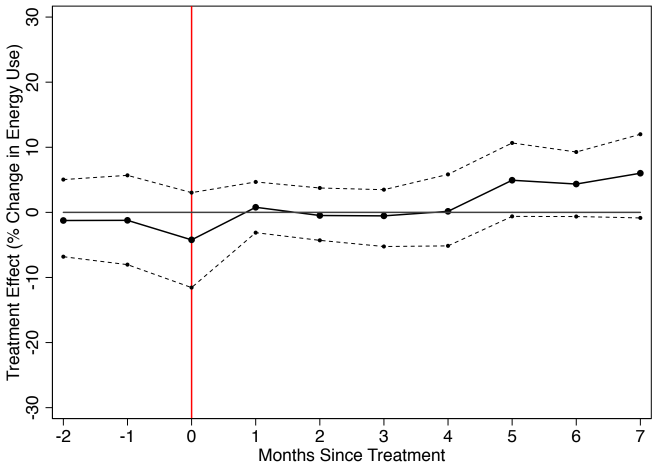


Figure 7: Time Varying Treatment Effects by Level of Energy Use

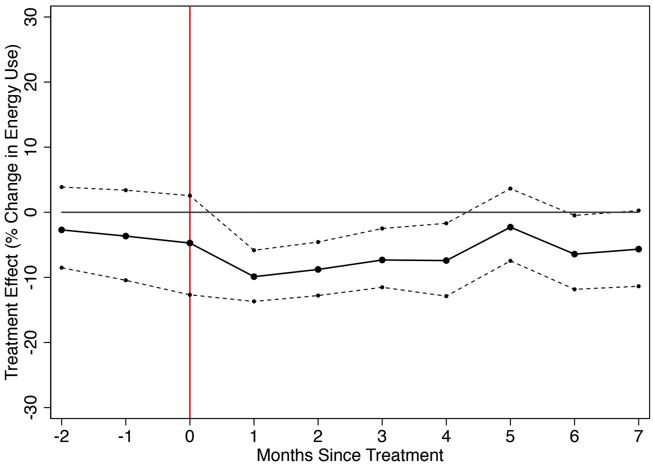
(a) Low users (0–20% quintile)



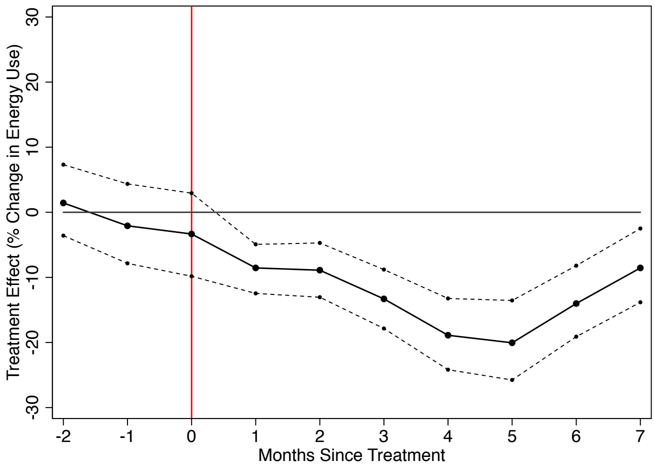
(b) Below average users (20–40% quintile)



(c) Above average usesr (40–60% quintile)



(d) High users (60–80% quintile)



For Online Publication

Supplemental Figures

Supplemental Tables

Figure A.1: Distribution of Average Daily Pre-Treatment Energy Use for Survey Respondents and Non-Respondents

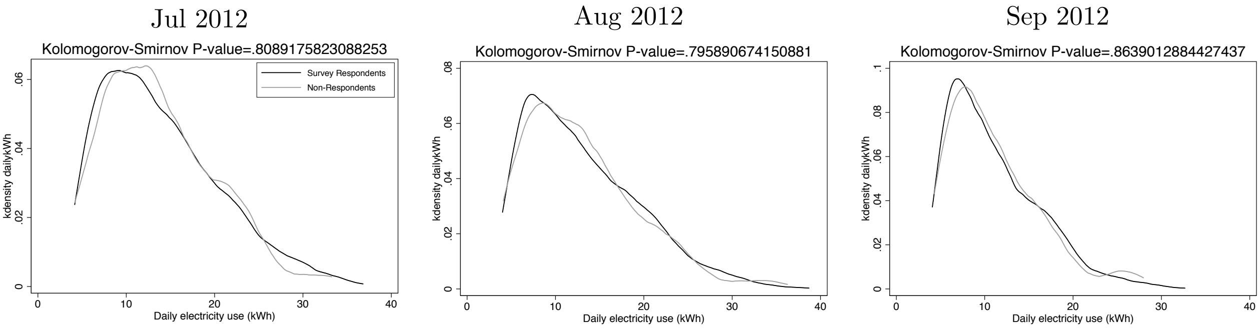


Figure A.2: Distribution of Monthly Pre-Treatment Energy Use for Survey Respondents and Non-Respondents

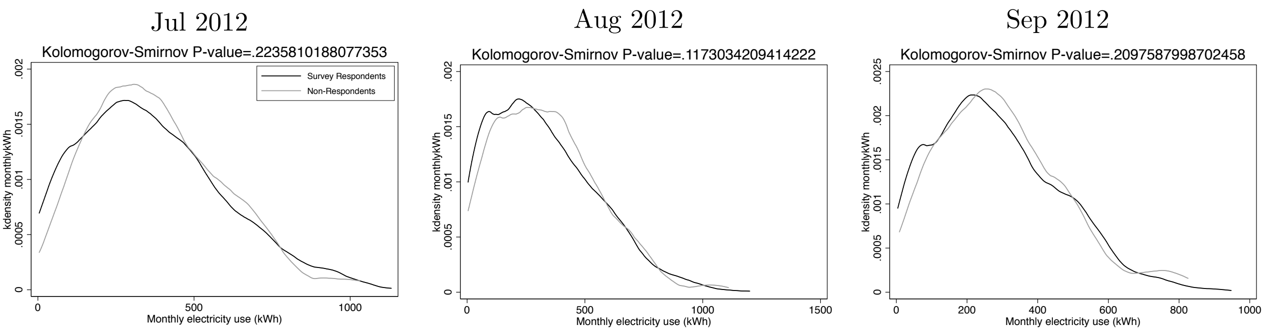


Figure A.3: Distribution of Average Daily Pre-Treatment Energy Use for Attritors and Non-Attritors

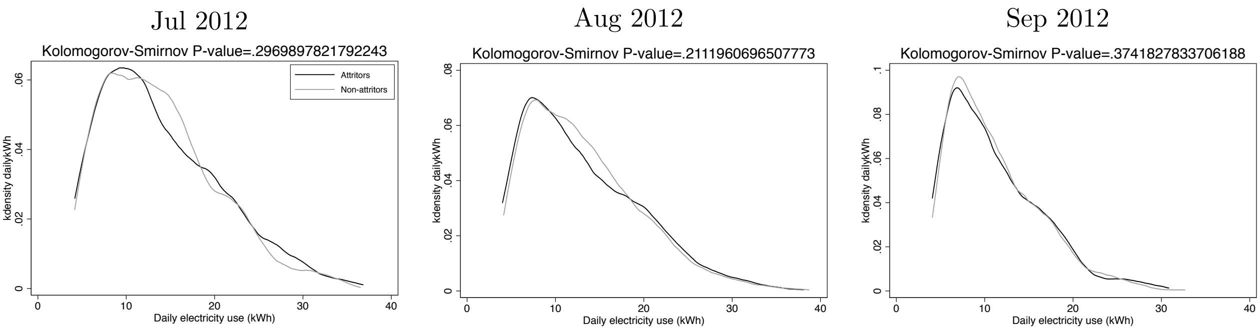


Figure A.4: Distribution of Monthly Pre-Treatment Energy Use for Attritors and Non-Attritors

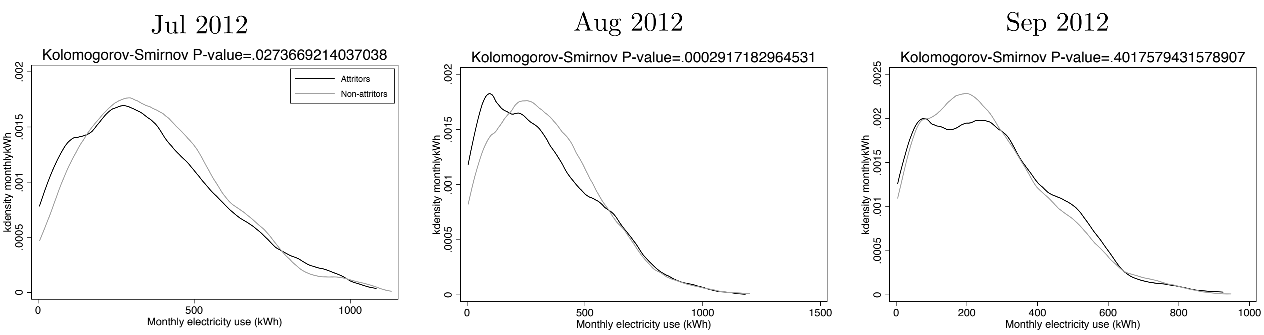
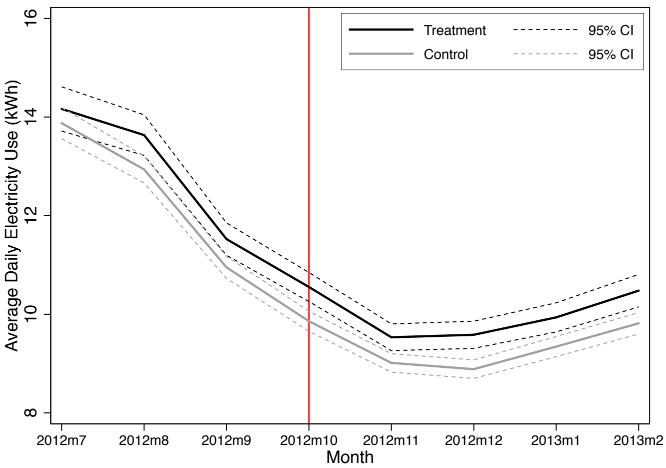
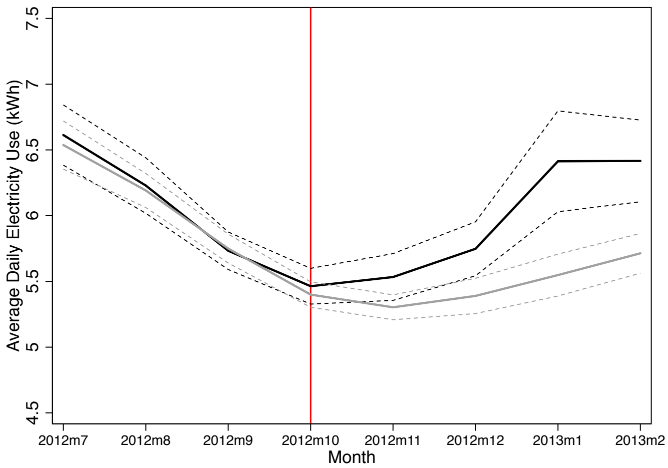


Figure A.5: Average Daily Energy Use by Use Level in Levels

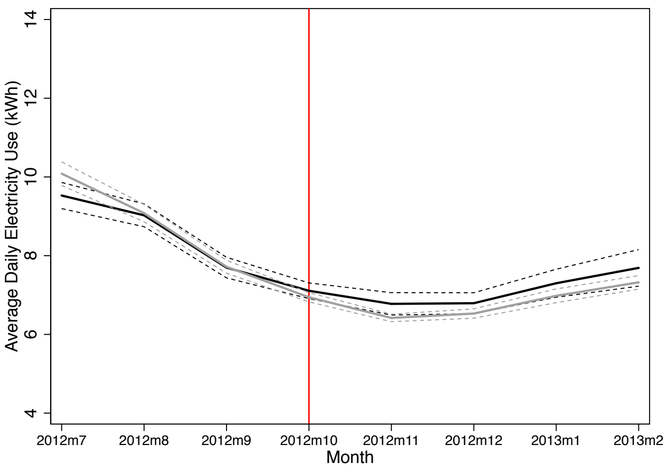
(a) Entire sample



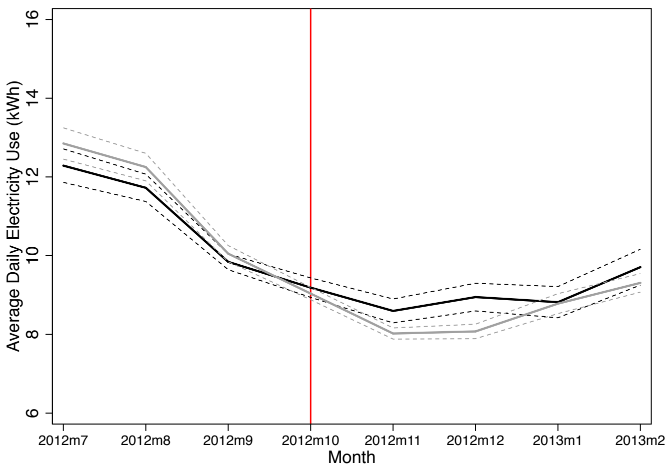
(b) Low users: 0–20% quintile



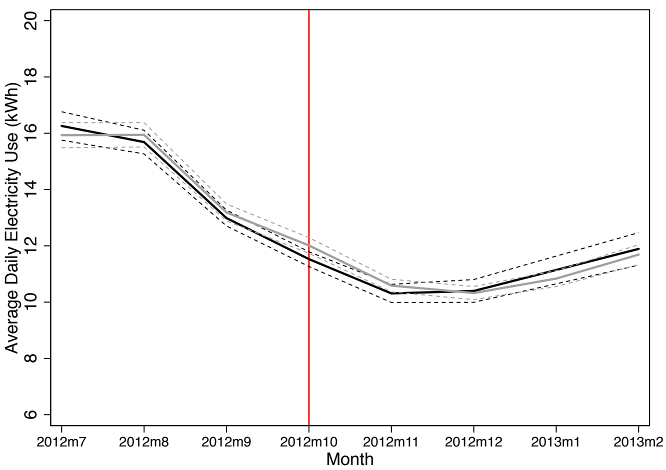
(c) Below average users: 20–40% quintile



(d) Average users: 40–60% quintile



(e) Above average users: 60–80% quintile



(f) High users: 80–100% quintile

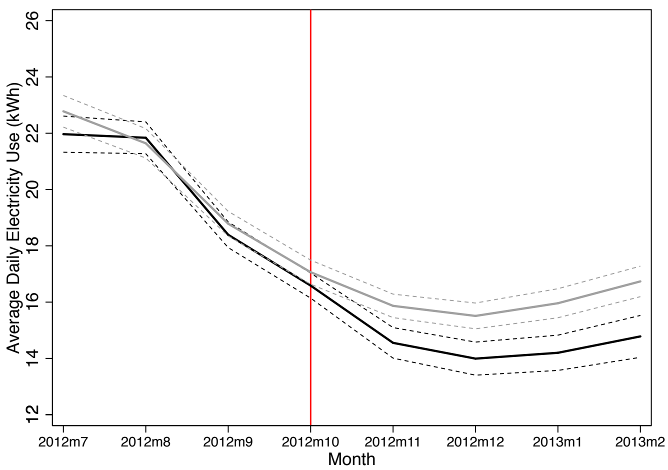
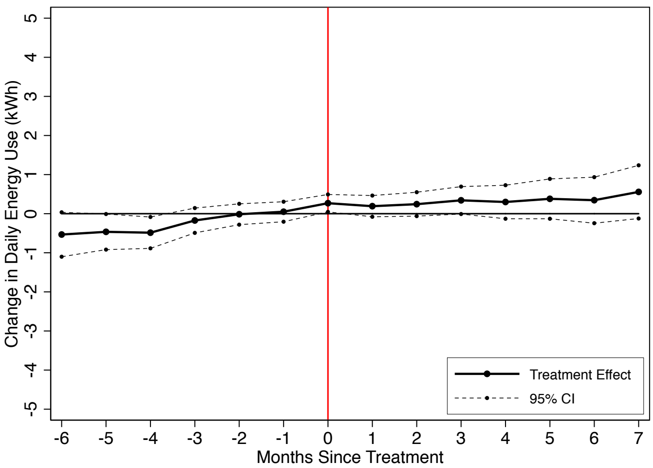
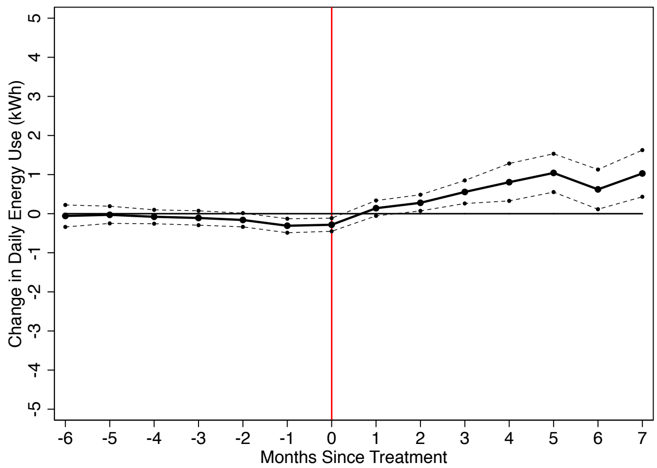


Figure A.6: Difference in Average Daily Energy Use between Treatment and Control Within Electricity Use Quintile (Extended Pre-Treatment Period Plots)

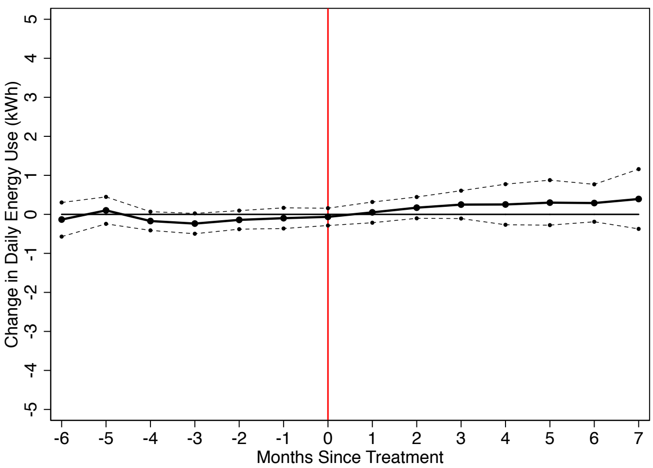
(a) Entire sample



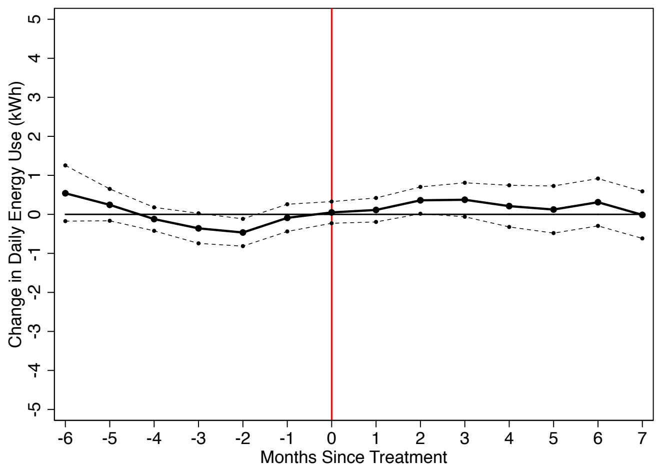
(b) Low users: 0–20% quintile



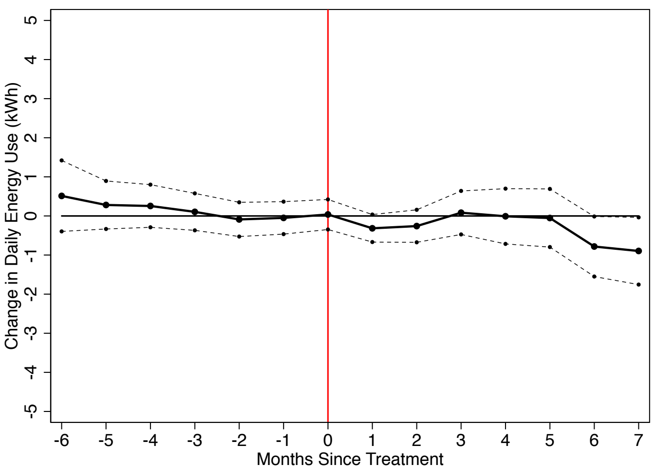
(c) Below average users: 20–40% quintile



(d) Average users: 40–60% quintile



(e) Above average users: 60–80% quintile



(f) High users: 80–100% quintile

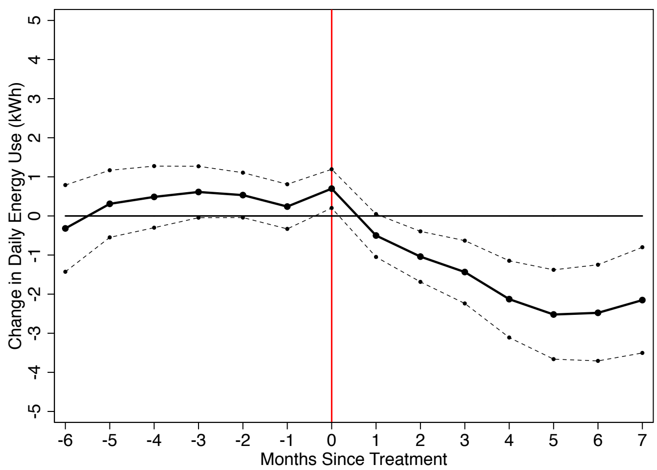


Table A.1: Pre-Treatment Monthly Energy Use (kWh): Survey Respondents and Non-Respondents

	Survey respondents	Non-respondents	Diff.
July 2012	379.27	364.40	14.87
			(14.92)
Aug 2012	338.50	325.90	12.60
			(13.75)
Sep 2012	296.71	283.45	13.26
			(11.28)
Number of households	311	2112	

Notes: Standard errors of differences in means are in parentheses and are clustered at the household level. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$.

Table A.2: Summary Statistics by Survey Response: Full Sample of Survey Respondents and Non-Respondents

	Survey respondents	Non respondents	Diff.
Census data			
Median household income	1377.16	1331.96	45.20
			(23.96)
Average age	37.16	36.79	0.37
			(0.28)
Full-time employment rate	42.41	41.01	1.40
			(0.57)
Proportion of households renting	37.54	38.84	-1.29
			(0.90)
Above median vote share for Green Party	0.49	0.49	0.00
			(0.02)
Number of households	1188	7376	

Notes: Population of residential households without solar panels whose energy use does not exceed 50 kWh/day. All variables that start with "Has" are dummies that equal one if the answer is yes and zero otherwise. Means and differences in means between treatment waves reported. Standard errors of differences in means are in parentheses and are clustered at the household level. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$.

Test of rank preservation assumption from Bitler et al. 2008 and Djebbari and Smith 2008

The following table produces the test of rank preservation from Bitler et al. 2008 and Djebbari and Smith 2008. Recall that if rank preservation holds, then each quantile of the control outcome distribution corresponds to the counterfactual outcome in the treatment distribution, and Quantile Treatment Effects estimates such as those in Figure 5 for March 2013 can be interpreted as the distribution of treatment effects.

The test of rank preservation follows the intuition that if ranks are preserved in the outcome distribution of the treatment and control groups, then exogenous covariates should have the same distributions within each quantile of the outcome distribution. The following table reports evidence against this. The table contains differences means between treatment and control households within four quantiles of the outcome distribution (daily electricity use) for March 2013. We report differences in means for quartiles of the outcome distribution, as per Djebbari and Smith 2008. For instance, the upper left corner indicates that weekly median income has a statistically significant larger mean in the 0–25^{th} among treatment households within the $0 - -25^{th}$ quantile of the outcome distribution in March 2013.

In seven of the twenty tests, we find a statistically significant difference in means of covariates between treatment and control, which is evidence against the rank preservation assumption.

Table A.3: Treatment and Control Differences at Quantiles of the Outcome Distribution

	0–25^{th} percentile	25–50^{th} percentile	50–75^{th} percentile	75–100^{th} percentile
Median weekly income	184.96**	117.20	42.01	15.86
Average age	2.89***	1.61*	1.21*	0.95
Full-time employment rate	2.33	3.61*	-0.78	2.07
Proportion of home renters	-6.11*	-9.70***	-4.38*	-1.36
Has above median vote share for Green Party	0.01	-0.19***	-0.08	-0.02

Notes: The outcome variable the quantiles are constructed with is monthly electricity use. Standard errors of differences in means at each quantile of the outcome distribution are clustered at the SA1 level. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table A.4: IV Estimates of Local Average Treatment Effects (LATE) of Home Energy Reports on Energy Use by Informedness, Baseline Energy Use and Demographics

	Dependent variable: log (mean daily kWh)	
	(1)	(2)
Treatment	1.43	7.08***
	(1.84)	(2.31)
(Underestimate relative energy use)	2.08	3.25
	(4.89)	(4.87)
(Correct about relative energy use)	-0.28	1.24
	(6.61)	(6.40)
(Overestimate relative energy use)	6.37	6.97*
	(3.91)	(3.90)
(Low energy use: 1^{st} quintile)	15.18***	16.50***
	(2.97)	(3.17)
(Below avg energy use: 2^{nd} quintile)	2.50	4.43
	(2.69)	(2.86)
(Above avg energy use: 4^{th} quintile)	-10.22***	-8.80***
	(3.31)	(3.43)
(High energy use: 5^{th} quintile)	-17.41***	-15.95***
	(3.61)	(3.98)
(High income)		-2.87
		(2.80)
(High age)		-2.91
		(2.34)
(High employment rate)		-3.12
		(2.84)
(High home rental rate)		-7.34***
		(2.55)
(High Green Party support)		18.22
		(13.64)
Heterogeneous post-treatment trends in electricity use		
Informedness	Y	Y
Pre-treatment use	Y	Y
Census (SA1) characteristics N	Y	
R^2	0.23	0.23
Observations	20129	20129
Treatment households	1531	1531
Control households	892	892

Notes: All specifications include household and month-of-year fixed effects and are weighted by attrition-correction IPWs. Standard errors are clustered at the household level using a pairs cluster bootstrap that accounts for estimation error in implementing IPWs. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$. The treatment variable is a dummy variable equaling one if a household accesses information from the smart-meter web portal. All variables in brackets correspond to interactions of that variable with the treatment dummy variable. The dummy variables for the energy-use quintiles correspond to baseline (pre-treatment) energy-use quintiles. High Income is a dummy that equals one if a household lives in an SA1 location whose average income is above median SA1 average income for all SA1s in the sample. The other "High" demographic variables are similarly defined; see the text for details.

Table A.5: Treatment Effects of Home Energy Reports on Energy Use by Informedness, Baseline Energy Use, and Demographics – Data Not Reweighted using Inverse Probability Weights

	Dependent variable: log (mean daily kWh)				
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.14	-0.24	0.68	5.46***	5.30***
	(0.95)	(0.99)	(1.35)	(1.54)	(1.72)
(Underestimate relative energy use)		-6.88	-0.28	-5.51	0.22
		(3.83)	(3.95)	(3.79)	(3.88)
(Correct about relative energy use)		-1.70	-1.07	-0.27	-0.24
		(5.27)	(5.11)	(5.00)	(4.87)
(Overestimate relative energy use)		11.62***	6.15**	12.57***	6.48***
		(2.68)	(2.71)	(2.71)	(2.71)
(Low energy use: 1^{st} quintile)			10.38***		11.09***
			(1.95)		(2.07)
(Below avg energy use: 2^{nd} quintile)			1.78		2.69
			(2.07)		(2.13)
(Above avg energy use: 4^{th} quintile)			-6.34***		-5.66***
			(2.25)		(2.37)
(High energy use: 5^{th} quintile)			-11.51***		-10.79***
			(2.40)		(2.58)
(High income)				-2.60	-2.66
				(1.92)	(1.85)
(High age)				-2.62*	-1.84
				(1.58)	(1.64)
(High employment rate)				-3.67*	-1.64
				(2.07)	(2.00)
(High home rental rate)				-4.59***	-5.53***
				(1.82)	(1.78)
(High Green Party support)				11.75	11.38
				(8.81)	(8.16)
Heterogeneous post-treatment trends in electricity use					
Informedness	N	Y	Y	Y	Y
Pre-treatment use	N	N	Y	N	Y
Census (SA1) characteristics	N	N	N	Y	Y
R^2	0.22	0.22	0.24	0.23	0.24
Observations	20129	20129	20129	20129	20129
Treatment households	1531	1531	1531	1531	1531
Control households	892	892	892	892	892

Notes: All specifications include household and month-of-year fixed effects. Standard errors are clustered at the household level. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$. The treatment variable in columns (1)–(5) is a dummy variable equalling one if a household is offered access to the smart-meter web portal. All variables in brackets correspond to interactions of that variable with the treatment dummy variable. The dummy variables for the energy-use quintiles correspond to baseline (pre-treatment) energy-use quintiles. High Income is a dummy that equals one if a household lives in an SA1 location whose average income is above median SA1 average income for all SA1s in the sample. The other "High" demographic variables are similarly defined; see the text for details.